PB-244 616

HAZE FORMATION: ITS NATURE AND ORIGIN

David F. Miller, et al

Battelle Columbus Laboratories

Prepared for:

Coordination Research Council, Incorporated
Environmental Protection Agency

March 1975

DISTRIBUTED BY:

NTIS

National Technical Information Service
U. S. DEPARTMENT OF COMMERCE

254154

PB 244 616

March, 1975

Final Report

HAZE FORMATION:
ITS NATURE AND ORIGIN

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. Department of Commerce
Springfield, VA. 22151

Battelle
Columbus Laboratories

A report of research conducted for the
Coordinating Research Council Inc.
and
U.S. Environmental Protection Agency

BIBLIOGRAPHIC DATA SHEET	1. Report No. CRC-APRAC-CAPA-6-68-5	2.	3. Recipient's Accession No.
4. Title and Subtitle Haze Formation: Its Nature and Origin			5. Report Date March 1975 6.
7. Author(s) D. F. Miller, Warren E. Schwarts, J. L. Gemma, Arthur Levy			8. Performing Organization Rept. No.
9. Performing Organization Name and Address Battelle, Columbus Laboratories 505 King Avenue Columbus, Ohio 43201			10. Project/Task/Work Unit No. 11. Contract/Grant No.
12. Sponsoring Organization Name and Address U.S. Environmental Protection Agency, Research Triangle Park, N. C. 27711 and the Coordinating Research Council, Inc., 30 Rockefeller Plaza, New York, N.Y. 10020			13. Type of Report & Period Covered Final 14.

15. Supplementary Notes

Not releasable to the public until August 1975.

16. Abstracts

This report documents the final year of a three year study of haze formation. The report is primarily concerned with the analysis of field data collected in previous years. Computer programs in time-series analysis were used to develop empirical models relating hourly light scattering averages to other air quality and meteorological data. The models, constructed from a continuous period of aerometric data in New York City, contain CO, NO_2, relative humidity, and total hydrocarbons as predictor variables. The models have been successfully applied to quantitatively predict light scattering in other regions.

17. Key Words and Document Analysis. 17a. Descriptors

Haze
Aerosols
Light Scattering
Particulate Matter
Air Pollution
Inorganic Composition
Organic Composition
Carbon Monoxide
Hydrocarbons
Humidity
Nitrogen Dioxide

17b. Identifiers/Open-Ended Terms

New York City, New York

PRICES SUBJECT TO CHANGE

17c. COSATI Field/Group 04B; 07B; 07C; 07D; 11D; 13B; 20N.

18. Availability Statement	19. Security Class (This Report)	21. No. of Pages
Releasable to the public.	UNCLASSIFIED	103
	20. Security Class (This Page) UNCLASSIFIED	22. Price 5.25/2.25

FORM NTIS-35 (REV. 3-72) THIS FORM MAY BE REPRODUCED USCOMM-DC 14952-P72

FINAL REPORT

HAZE FORMATION: ITS NATURE AND ORIGIN

to

COORDINATING RESEARCH COUNCIL INC.
and
U.S. ENVIRONMENTAL PROTECTION AGENCY

March, 1975

by

David F. Miller, Warren E. Schwartz,
James L. Gemma, and Arthur Levy

BATTELLE
Columbus Laboratories
505 King Avenue
Columbus, Ohio 43201

ACKNOWLEDGMENTS

We would like to acknowledge the help and guidance of the people who served on the CAPA 6-68 Project Committee:

Mr. B. S. Bailey – Chairman
Dr. R. L. Bradow
Dr. M. K. Gest
Dr. L. C. Gibbons
Dr. T. P. Goldstein
Dr. P. J. Groblicki
Mr. W. Lonneman
Dr. H. W. Otto
Mr. R. Patterson
Dr. T. R. Powers
Mr. J. W. Shiller
Dr. J. Vardi
Dr. J. Wagman
Dr. E. E. Weaver
Dr. F. T. Weiss
Mr. A. E. Zengel

TABLE OF CONTENTS

	Page
INTRODUCTION	1
BACKGROUND	2
Characteristics of Haze	2
Chemical Characteristics of Aerosols	5
SCOPE OF PROGRAM	6
First Year	6
Second Year	7
Third Year	7
MAJOR FINDINGS AND CONCLUSIONS	8
Light-Scattering Interpretations	8
Features of Aerosol Composition	10
RESULTS AND DISCUSSION	12
Light-Scattering Measurements	12
Statistical Analysis of Light Scattering	12
Preliminary Analyses	14
Time-Series Analysis	15
Forecasting	16
Summary	19
Chemical Composition of Light-Scattering Aerosols	35
Inorganic Composition	35
Organic Composition	41
Solvent-Extractable Particulate Matter	44
Weight Percent C, H, and N	46
Infrared Spectra	46
Aromatic/Aliphatic Ratio	53
Acid/Base Neutral Distribution	53
Functional Group Analyses	55
Summary	57
REFERENCES	61

APPENDIXES

A. STATISTICAL METHODS	A-1
B. ORGANIC ANALYTICAL PROCEDURES	B-1
C. INFRARED SPECTRA OF ORGANIC PARTICULATE	C-1

FINAL REPORT

on

HAZE FORMATION: ITS NATURE AND ORIGIN

to

COORDINATING RESEARCH COUNCIL INC.
(CAPA 6-68)

and

U.S. ENVIRONMENTAL PROTECTION AGENCY
(Contract No. 68-02-0792)

from

BATTELLE
Columbus Laboratories

March, 1975

INTRODUCTION

In 1970, when this program was undertaken by Battelle-Columbus for the Coordinating Research Council and the Environmental Protection Agency, very little was known regarding the nature of aerosols responsible for haze – a condition which accompanies other manifestations of smog and often serves as a visual indicator of the severity of smog. About that time an important aerosol study was conducted in the Pasadena area in which nearly all available tools were utilized in an intensive monitoring and analytical effort to characterize the airborne particulate in the Los Angeles Basin. In spite of the breadth of that project, many gaps remained in our knowledge, particularly regarding the chemical composition of aerosols and the predominant factors influencing aerosol formation. Frustrated by the initial goal of developing a direct means of determining the total automotive contribution to haze in urban areas, the subject program was redirected to investigate more generally the aerosols responsible for haze with emphasis placed on organic analyses and analyses of the factors influencing light scattering. Although we feel that the results of this program have significantly advanced our overall understanding in these areas of aerosol science, analytical and interpretive difficulties still abound, and a thorough accounting of the nature and origins of haze requires continued investigations.

The program was conducted as a 3-year study beginning in June, 1970, and ending December, 1974. There was a 1-year interruption in the work between the first-year and second-year programs. Reports summarizing the progress during the first 2 years were issued separately in 1972[1] and 1973[2]. This report covers the third year of progress, and for convenience of discussion, includes summaries of data from the Second Year Report. The principal sections of the report cover results on statistical analyses of light-scattering data and chemical analyses of light-scattering aerosols. Preceding these sections we have included short sections covering the phenomena of haze, some related research results, the scope of the program, and the major findings of the study.

BACKGROUND

Characteristics of Haze

Haze, by definition, is a quality of indistinctness – in the context of atmospheric science it refers to a blurring or reduction in the perception of distant objects. The visual effect of haze is similar to that of fog, although a color effect is more frequently observed with haze. In contrast with haze, fog consists principally of water droplets, although it is well established that foreign particles (condensation nuclei) are usually required to nucleate water vapor in advance of condensation to fog. Aerosols associated with haze may also contain water but are comprised primarily of solids and other liquid materials which have either been emitted into the atmosphere from a variety of sources, or produced in the atmosphere as a result of photoinduced reactions of gaseous pollutants. The latter category of aerosols is commonly referred to as secondary or photochemical aerosols.

Visibility problems associated with haze have been recognized for some time. The significance of haze in comparison with fog is shown in Figure 1 in which the yearly occurrences of days with visibility <7 miles in New York City is plotted from 1934 to 1960.[3] The number of poor visibility days attributable to fog has remained fairly constant while the occurrence of haze has increased dramatically since 1946.

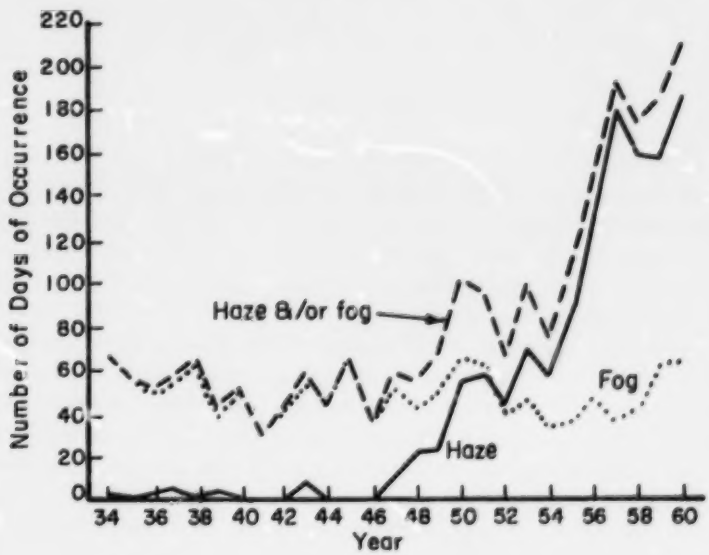


FIGURE 1. FREQUENCY OF POOR VISIBILITY DAYS IN NEW YORK CITY ATTRIBUTABLE TO HAZE AND FOG[3]

Evaluation of visibility is most frequently performed by human observation, and, aside from the limitations of subjective measurements, visual quality has some inherent complexities. Among the complications is the fact that particles scatter more light in a forward direction than in a backward direction. Thus the same aerosol cloud may appear more dense when viewed toward the sun than away from the sun. As alternatives to direct observations, light-transmission instruments (transmissometers) are commonly used for estimating visual quality. The instruments

measure the extinction of visible light over a large distance which must in turn be related to observed visibilities. In most cases the Koschmeider Formula (visual range = 3.9/extinction coefficient) is used for such calculations. Assumptions inherent in the formula and conditions affecting its application have been reviewed.[4] Light extinction, as measured directly or by inference from such instruments, is necessarily an average value over a considerable distance in which many inhomogeneities may exist. The shortcoming of such determinations is that they cannot be readily related to other air-quality parameters determined at point sources.

In this program, haze or visual quality was estimated using an integrating nephelometer — a method conceived by Beuttell and Brewer[5] and later modified for air monitoring by Ahlquist and Charlson[6]. The instrument measures the volume scattering coefficient of a discrete air sample pulled through an optical chamber. The volume scattering coefficient (b_{scat}) is determined by integration of the scattering over nearly half a solid angle.

It should be noted thet the nephelometer measures the extinction of light due to scattering alone while the total extinction of light along an atmospheric path occurs by both scattering and absorption of gases and aerosols. Rayleigh scattering, which dominates in clean air, occurs from air molecules and particles approaching molecular size. Mie scattering is important for particles whose dimensions are roughly comparable to the wavelength of visible light. Thus aerosols in the size range 0.1 to 1-μm diameter are extremely effective Mie scatterers. Absorption of light by gases and particles may be significant in polluted atmospheres, but, except for unusual cases, Mie scattering by aerosols largely dominates the effect of visibility degradation. Rayleigh scattering as determined by the integrating nephelometer is relatively constant and corresponds to a b_{scat} value of $0.2 \times 10^{-4} m^{-1}$ (b_{scat} units of $10^{-4} m^{-1}$ are commonly used with the nephelometer).

Many investigations have been conducted to associate nephelometry data with visibility and aerosol mass-loading data. While the relationships of b_{scat} with total particulate mass concentrations are not always good, as might be expected, the relationships with visibility are generally quite good. In a recent study conducted in the Los Angeles, Oakland and Sacramento areas, the correlation coefficients between b_{scat} and prevailing visibility* ranged from 0.84 to 0.91 for the three sites.[7]

For convenience in relating b_{scat} values to visual range, a scale is provided in Figure 2 based on the empirical relationships developed by Charlson for the nephelometer.[8] Light scattering values for clean air and Freon 12 calibration are indicated.

Aside from the studies relating light scattering to aerosol concentrations and visibility, few studies have concentrated on interpreting light-scattering data in relation to other air-pollution phenomena. In some early work by Buchan and Charlson in Seattle[9], significant correlations were established between 10-minute averages of NO_x concentrations and the nephelometer light-scattering coefficient. In more recent studies by Covert et al.[10], growth related to deliquescent properties of simulated aerosols was demonstrated, and humidity effects observed while monitoring light scattering in the coastal regions of California were attributed to this behavior. Lundgren[11] conducted a study in which size-classified aerosols were collected on a 4-hour basis to assess diurnal variations in particulate composition and concentration. Results of the study indicated that light scattering correlates well with particulate nitrate, peroxyacetyl nitrate (PAN) and total particulate concentration when windy days are omitted from the analysis.

*Prevailing visibility is defined as the greatest visibility which is attained or surpassed around at least half of the horizon circle. Visual range is the distance at which it is just possible to visualize with the unaided eye the contrast of a prominent dark object against the daytime horizon sky.

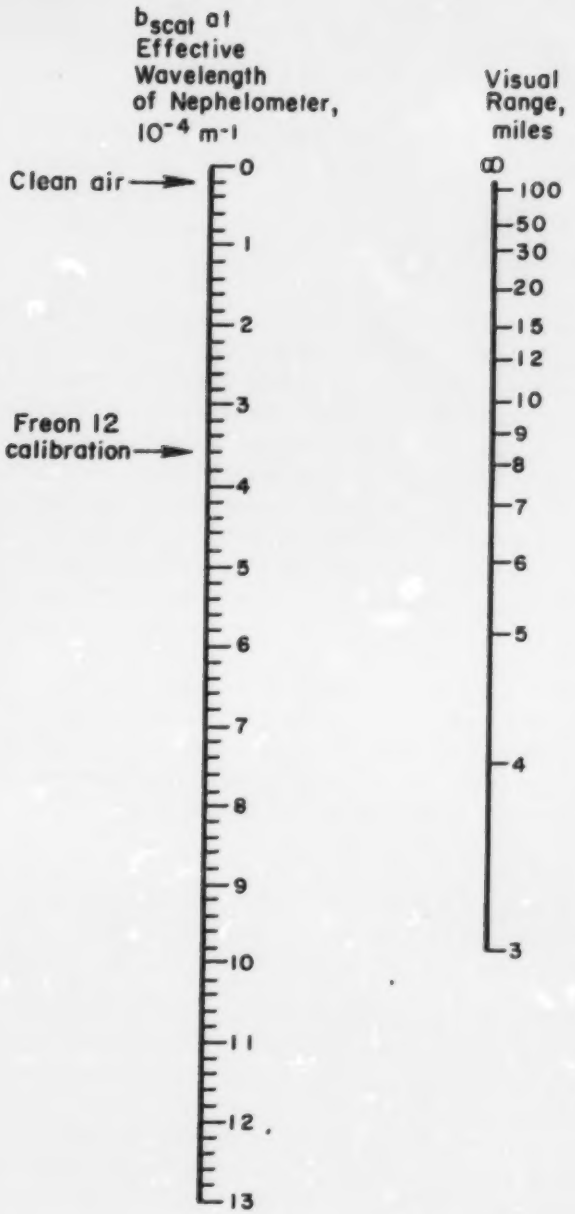


FIGURE 2. RELATIONSHIP BETWEEN VISUAL RANGE AND LIGHT SCATTERING MEASURED BY AN INTEGRATING NEPHELOMETER[8]

Results of the 1969 Pasadena study[12] showed that the most pronounced increases in light scattering and aerosol volume are coincident with maximum solar radiation intensity. Analyses of data from the 1972 Air Resources Board project in California indicated that the relationship of light scattering with ozone is not well established, and autocorrelation results show complex relationships for light scattering that are not well understood.[13]

Chemical Characteristics of Aerosols

Numerous studies have been conducted in an effort to characterize the chemical composition of aerosols. Most of the work has been on total suspended particulates, and many analyses have dealt only with obtaining the distribution of chemical elements in the particulates. A comprehensive elemental analysis of "hi-vol" samples from six CAMP stations (Cincinnati, St. Louis, Washington, D.C., Denver, Philadelphia and Chicago) was reported by Blosser.[14] John et al.[15] reported on trace-element concentrations of "hi-vol" particulate collected in San Francisco. More important to our understanding of the nature of haze have been studies involving the chemical analysis of aerosols according to size. Elemental analyses of cascade-impactor collections have been made by several investigators.[16-18] The size distributions of phosphate, nitrate, chloride, and ammonium[19] and that of sulfate[20] have also been studied using cascade impactors. Some of the aforementioned studies, and particularly studies conducted by Friedlander and his associates[21-23], have been directed at identifying sources of aerosols on the basis of emissions data and elemental tracers of certain origin. The overall results of most of this work show that in polluted atmospheres most of the naturally occurring aerosols are in the >2 μm-size range together with emitted and/or reentrained matter such as fly ash, tire dust, and cement dust.

There appears to be general agreement that the aerosols <2 μm are composed predominantly of sulfur and organic compounds, with nitrogen compounds and some metals associated with combustion (Pb, Zn, and V) making minor contributions.

An interesting study by Novakov et al.[24] described diurnal variations in the chemical states of sulfur and nitrogen particulate in Los Angeles. The S^{6+} (sulfate) to S^{4+} (sulfite) ratio for particles <2 μm was shown to be >1 for only a few early-morning hours. Throughout most of the day the sulfate/sulfite ratio was near 0.5.

Far less information is available on the nature and origin of the organic matter of small aerosols, owing primarily to difficulties of organic micro analyses and the fact that the organic constituency does not contain compounds traceable to specific sources. While few detailed analyses have been performed it has become routine to determine the fraction of total particulate which is extractable by organic solvents, most often benzene. The benzene fraction is reported to range from 4-14 weight percent over a range of U.S. cities.[25] The aliphatic fraction of benzene-soluble matter has been studied by gas chromatography.[26] In a recent study by Ciaccio et al.[27], a variety of extracting solvents and spectroscopic methods were employed to analyze the organic fraction of aerosols collected along a highway complex in New York City. Their analyses, however, were performed on aerosols in the size range 0.4-7 μm, and they may be only marginally relevant to light-scattering aerosols. One of the more pertinent studies in organic analysis was conducted some 20 years ago by Mader et al.[28] Analyzing organic fractions of Los Angeles aerosols in the light-scattering range (0.3-0.8 μm), they observed sizeable amounts of oxygenated and peroxidic organic materials, and they demonstrated similarities between the absorption spectra of the Los Angeles aerosols and those generated synthetically upon irradiating gasoline hydrocarbons and NO_x in a smog chamber.

SCOPE OF PROGRAM

First Year

The initial objective of this program was to develop a method of distinguishing between light-scattering aerosols attributable to automotive sources as opposed to those from nonautomotive sources. Both chemical and optical approaches were considered initially, but upon reviewing the possibilities in more detail, emphasis was placed on chemical methods of analysis. Thus the first year of the program was devoted to seeking a chemical basis which would permit differentiation among the organic composition of aerosols according to precursors. For this purpose aerosols associated with the following sources were collected for analysis:

(1) Rural atmospheres, Blue Ridge Mountains, N.C.

(2) Controlled atmospheres, Battelle-Columbus smog chamber

(3) Diluted primary auto exhaust

(4) Urban atmospheres, Bronx, New York.

Chemical fractionation of particulate extracts was performed after which organic constituents were analyzed by gas chromatography combined with mass spectrometry. It became apparent that spectra of organic material present from each of the above sources is enormously complex. This is true even when a single hydrocarbon and NO_x are irradiated to produce aerosols.

Irradiation of α-pinene (one of several terpenes emitted in abundance by trees and plants) and nitrogen oxides in a smog chamber produced aerosols having chemical properties similar to that of aerosols collected in a forested region. In this case, pinonic acid was identified as a product of both the natural-rural (Blue Ridge Mountains) and the simulated-rural (smog chamber) aerosols. This finding served to confirm the utility of the smog chamber in simulating atmospheric conditions conducive to secondary aerosol formation.

Analyses of primary auto-exhaust aerosols revealed two predominant compounds, namely benzoic and phenylacetic acid, among the complex acidic fraction of the aerosol matter. However, these aromatic acids were not detected in urban particulate similarly analyzed. This result is rationalized in terms of dilution of auto exhaust in the atmosphere, or on the basis of removal of these acids from the atmosphere by photochemical reactions. Indeed, decarboxylation of organic acids was found to occur during the progress of aerosol formation under smog-chamber conditions.

The principal mission during the first year was to determine if a chemical basis could be established for estimating the atmospheric burden of secondary aerosols related to automotive emissions. Combined gas chromatography – mass spectrometry analyses revealed that the organic constituency of atmospheric aerosols was enormously complex – so complex that complete resolution was a formidable task. It was also obvious that automobile exhausts do not contain hydrocarbons uniquely related to automotive operations*. Thus at the conclusion of the first year, the approach of establishing unique precursor relationships between automotive emissions and the chemical composition of aerosols was abandoned, in spite of the evidence that aerosol precursor relationships could be established in smog-chamber simulations of simple systems. The problem as it related to auto exhaust was too complex, and too little was known about the chemical nature of aerosols in the atmosphere.

*There are a few hydrocarbons, such as acetylene and ethylene, which might be regarded as unique to automotive emissions on a relative basis, but these hydrocarbons do not participate in the formation of organic matter condensing into aerosols.

Second Year

At the beginning of the second year, a somewhat broader objective was defined which sought to improve on the understanding of the composition and sources of aerosols contributing to urban haze. The overall objective was to determine the relationships which might exist between the concentration and composition of the aerosols responsible for reduced visibility and the sources and conditions under which these aerosols form in the atmosphere. In seeking these relationships, it was apparent that the customary methods of characterizing particulate composition were incomplete, particularly with respect to the organic constituency. Secondly, it was obvious that a complete analysis of the air was necessary concomitant with the aerosol collections. Thus the second year's program had two specific objectives (1) to develop and apply an analytical scheme for characterizing the broad chemical classes and functional groups present in the organic fraction of light-scattering aerosols and (2) to conduct field-sampling programs for collecting aerosols and characterizing the prevailing atmosphere conditions under which the aerosols accumulated.

Field sampling was conducted in downtown Columbus during the latter half of July, 1972, in New York City (Welfare Island) throughout August, and in Pomona, California, for 10 days in mid-November. The sampling involved continuous monitoring of the meteorological conditions, integrated light scattering (visibility), solar radiation intensity and the gas-phase composition of the air (including CO, NO, NO_2, SO_2, ozone, methane, ethylene, acetylene, and total hydrocarbon) while simultaneously collecting aerosols for chemical analyses. All aerosol samples were collected on a diurnal basis. To provide the aerosol mass needed for organic analyses, three samplers operated continuously at 20 cfm, sampling 25 feet above ground level. Because the program is concerned with small particles which cause light scattering, the samplers were equipped with size-fractionating devices which provided particle separation (based on aerodynamic size) near 2 μm diameter.

An analytical scheme was developed which provides semi-quantitative data on the organic features of aerosols. Diurnal differentiation of samples limited the quantities of material available for analyses. The organic matter extracted from the daily aerosol collections (<2 μm diameters) averaged approximately 14 mg per extraction. Within such limitations, both spectroscopic and wet-chemical analyses were conducted. The analyses provide numerical data on the weight-percent solvent extractable (methylene chloride and dioxane) components of the aerosols, weight-percent CHN in the extractables, infrared spectroscopic band intensities for specified absorptions, aromatic/aliphatic ratios and the concentration of total alcohol and total carbonyl in specified sample fractions. During the second year's program, six aerosol samples were analyzed; two samples from each of the three sampling sites. In addition, inorganic analyses were performed on aerosol samples selected from 22 of the 43 sampling days at the three sites. The analyses included trace metals by optical emission spectrometry, Cl, Br, S, and Pb by X-ray fluorescence, and $SO_4^=$, NO_3^-, and NH_4^+, and CHN by other chemical methods.

Third Year

To fulfill the overall objective set forth in the second year, two specific objectives were defined for the third year of the program: (1) to apply the organic analytical scheme to additional aerosol samples and to estimate the precision of the analytical procedure, and (2) to conduct a statistical analysis of the aerometric and aerosol compositional data with emphasis on the development of an empirical model of light scattering based on air pollution data.

Organic analysis were conducted on three atmospheric samples collected by the groups noted in parentheses: Denver, Col. 11-17-73 (EPA), West Covina, Ca. 9-21-73 (Battelle-Columbus), Rubidoux, Ca. 9-21-73 (EPA/California Air Resources Board). A week-long collection of aerosol in Pomona, Ca. (Oct., 1972) by EPA was composited to obtain enough aerosol mass for replicate analyses in estimating analytical accuracy. In addition, organic analyses were conducted on an aerosol sample representative of primary and primary plus secondary auto-exhaust aerosol. The latter sample was a composite of auto exhaust aerosol collected after irradiating diluted exhaust in Battelle's smog chamber.

Utilizing data from the 1972 field study in New York, an empirical model was developed relating light scattering to other measures of air quality. A computer program for multivariate classification was used to select the independent variables to be considered for model development. From the preliminary information, time-series models were created which could predict the degree of light scattering on the basis of changes in CO, NO_2, total hydrocarbons and relative humidity. Applications of the model to short-term forecasting of visibility in other localities (Denver, Col. and Dayton, Ohio) were successful. The modeling results were summarized in a paper entitled "A Model of Urban Visibility Based on Air Quality" presented at the 168th National Meeting of the American Chemical Society, Atlantic City, New Jersey, 1974.[29]

The procedures and results of the statistical analysis of ambient data, and those of the chemical analyses of aerosols, comprise the major portion of this report.

MAJOR FINDINGS AND CONCLUSIONS

Light-Scattering Interpretations

During the combined air-sampling period of more than 3 months in New York and Ohio, the maximum hourly-average light-scattering coefficient was 12 x $10^{-4}m^{-1}$, corresponding to a visual range of about 2 miles (Figure 2). It is noted that inferences here to visual range relate to single point measurements. During the same sampling period, the lowest hourly-average scattering coefficient was 0.9 x $10^{-4}m^{-1}$ (visibility of 35 miles), and the grand average value was 3.4 x $10^{-4}m^{-1}$ (visibility of 9 miles). In contrast, light-scattering measurements performed in the Blue Ridge Mountains during the season of blue haze indicated a maximum value of 2 x $10^{-4}m^{-1}$ (visibility of 15 miles) and a minimum value approaching 0.2 x $10^{-4}m^{-1}$((visibility >100 miles); a minimum value corresponding to the coefficient of molecular (Rayleigh) scattering of clean air. Unlike frequent conditions of maximum visibility in the Blue Ridge Mountains, the air in the urban areas was always significantly polluted with light-scattering particles.

The average diurnal pattern of light scattering in New York and the two Ohio cities (Columbus and Dayton) showed similar features. Typically, a steady increase in light scattering started in the early evening and continued until sunup. This trend is most likely attributable to surface inversions that begin at sunset.(30) The overall light-scattering patterns also revealed slight fluctuations coincident with increased automobile traffic during the morning and late afternoon rush hours. With few exceptions, noonday and early-afternoon maxima in light scattering were absent, thus leading to the conclusion that photochemical reactions are not the dominant processes involved in the generation of light-scattering aerosols in these regions of the country.

The hour-by-hour changes in light scattering in these cities can be described satisfactorily by an empirically derived time-series model which predicts light scattering on the basis of the daily trends of the scattering coefficient and other air-quality factors, namely the concentrations of nitrogen dioxide, carbon monoxide, total hydrocarbon and relative humidity. We have shown that the relatively simple time-series models developed here are capable of predicting the next hourly light-scattering average within 5-10 percent of the actual b_{scat} average at the 50-percent-confidence level. The model also proved useful in forecasting light scattering several hours in advance of existing air-quality data. For example, for 4-hour lead times, the model predicted light scattering within 10 and 17 percent of the actual values in Dayton and New York, respectively, at the 50-percent-confidence level. Forecasts of light scattering in hourly increments up to 8 hours were attempted, and the results are described in the text of the report.

It is important to note here that conclusions about causality cannot be drawn from the structure of the time-series models developed or from the variables comprising the models. The variables were chosen because, statistically, they form a set of predictors which could best explain the variance in light scattering. A variable such as CO, for example, is undoubtedly a "tracer" variable, that is not one which is directly involved in the process of haze formation, but one which in some way adequately summarizes information about other variables which are involved in the process.

The results of the modeling are very encouraging for the prospects of developing a somewhat more sophisticated and expanded model which would provide statistical predictions of afternoon and evening visibility based on trends in the atmospheric conditions of the morning and the previous day. Eventually, it might also be possible to couple the stochastic model to a kinetic chemical model to adequately explain the process of haze formation in smog.

As made evident throughout the report, there is a dramatic difference between the chemical composition of the light-scattering aerosols (i.e., diameters < 2 μm) and the larger aerosols (diameters > 2 μm). The finding supports the idea that the sources of aerosols in the two size ranges are decidedly different and that there is little interaction (e.g., agglomeration) between the aerosols in these size ranges. Coupling the rates of change observed in light scattering with the conclusions regarding particle interaction leads to other implications. Among them is the generalization that light-scattering aerosols in the regions sampled are continually and rapidly being diluted and removed by the influx of cleaner air. In general, coagulation of these aerosols which occurs with aging and which reduces the net scattering efficiency, and scavenging mechanisms which ultimately remove the aerosols from the air are both slow processes (half-lives of several days are typical) relative to the rate of dilution required to balance the apparent emission and/or formation rates of the light-scattering aerosols in these cities. (It was not uncommon to observe light scattering increasing at hourly-average rates of 1-2 b_{scat} units per hour, corresponding to aerosol-mass concentration increasing at rates of perhaps 40-80 $\mu g/m^3$ per hour.) Thus it appears that the Eastern cities are blessed by a consistent breeze which is particularly prevalent during the peak sunlight hours when solar energy is crucial to the development of photochemical aerosols. The overall result of the local cleansing is, of course, the transport of light-scattering aerosols to other regions of the country. And the net effect in the outlying regions is likely that of creating an artificial background of particles whose average concentration will be dominated by dispersion rates and scavenging rates. Superimposed on the "background" aerosol concentration will be the diurnal variations in concentration controlled by local emissions and local meteorological conditions.

Features of Aerosol Composition

Chemical analyses of aerosol samples collected in Columbus, New York, and Pomona show that there are vast differences in the chemistry of aerosols classified on the basis of aerodynamic diameters above and below 2 μm. The larger particles (> 2 μm) which play little part in the scattering of light are composed primarily of soil and sea compounds. Metallic compounds, silica, and carbonates predominate. Nitrate and sulfate compounds comprise a minor fraction of this matter.

In contrast, aerosols in the size range < 2 μm contain almost no metallic elements – lead and zinc are exceptions. They consist primarily of organic compounds and sulfur compounds. Sulfate accounted for about 20 percent of the small-particle mass, but sulfate sulfur accounted for only one-third of the total sulfur concentration. The concentration of sulfate was highly correlated with the ammonium concentration. Nitrate was a minor constituent of the aerosol in Columbus and New York but a substantial constituent in Pomona. Statistical analyses of the inorganic composition of the light-scattering aerosols in conjunction with the quantity of haze and other air-quality parameters revealed few statistically-significant relationships. Aside from the strong correlation between sulfate and ammonium, inorganic components are unrelated to each other. Considering air-quality factors, there was a significant positive correlation between the degree of light scattering and the percent concentration of sulfate and ammonium. A very high correlation exists between the percent lead in the aerosols and the daily average NO_2 concentration. For the most part, daily variations in the inorganic composition of aerosols are not large, and subtle relationships which might exist with atmospheric variations remain hidden by the complexities of the atmospheric conditions.

For organic analyses, samples of light-scattering aerosols (<2 μm) were subjected to soxhlet extraction using methylene chloride ($MeCl_2$), and some of the samples were subsequently subjected to a second extraction using dioxane. For 10 atmospheric aerosol samples, the weight percent $MeCl_2$ extractable averaged 17 percent, and ranged from 8 to 45 percent. For 7 aerosol samples subjected to a second extraction with dioxane, the weight percent dioxane extractable averaged

19 percent and ranged from 5 to 40 percent. In cases where both solvents were sequentially employed, total extractable matter averaged 33 percent, and ranged from 16 to 44 percent.

It was determined that solvent-extractable blanks for glass-fiber and quartz-tissue filters are significant compared to the mass of extractable material obtained from a daily collection, and preextraction of such filters is recommended in sampling for similar analytical purposes. The magnitude and variability of the dioxane blank was sufficiently large that data concerning weight-percent dioxane extractable matter should be interpreted cautiously. Owing to the uncertainties concerning the dioxane extracts most analyses were conducted using $MeCl_2$-extracted matter.

With the exception of the sample from the Trout farm area of Denver, which showed several peculiarities in composition, the weight-percent-extractable and the gross-elemental (i.e., the C, H, N, O distribution) compositions of the methylene chloride extracts of samples from New York, Ohio, and California were quite similar. These extracts were fractionated into water-soluble and water-insoluble components. The water-insoluble material was then fractionated into acid, neutral and basic component. The distribution of material among these classes served as part of the general characterization of the sample. In general, the $MeCl_2$ extracts of the West Coast samples contained a higher percentage of water-soluble matter than the East Coast samples (40 percent versus 15 percent). Among the water-insoluble fractions of these extracts, the neutral fractions account for 60-70 percent of the mass and the acid fractions account for 30-40 percent. Quantitative analyses for total carbonyl and total alcohol among the neutral fractions reveal the presence of about 5 weight percent oxygen for each of the functionalities. No consistent trends were apparent in the carbonyl and alcohol data, either with respect to their relative distributions or with respect to the total oxygen content of the aerosols or the concentration of ozone during the sampling periods.

Infrared analyses of the $MeCl_2$ extracts show the presence of a variety of carbonyl-containing compounds and possibly the presence of peroxide and/or carbonate compounds. There are also indications of small concentrations of organic nitrates, and the trends in the intensity of the nitrate bands correlate well with the total nitrogen content of the extracted aerosols. Fourier-transform-NMR spectrometry indicates that aliphatic protons are an order-of-magnitude more prevalent than aromatic protons.

The Denver (Trout farm) sample which was nearly one-half soluble in $MeCl_2$ contained relatively little oxygen and nitrogen, and appeared to be highly "saturated" compared to all the samples from other regions.

Excluding the unusual Denver sample, the overall characteristics of the organic composition of light-scattering aerosols are only moderately different for samples from different sites and are only moderately different on days of different air-quality at the same site. There appears to be a trend of higher percentages of total oxygenated matter on days of higher light scattering, but the more specific analyses for oxygen-bearing functionalities show no obvious patterns with the degree of light scattering or other air-quality parameters. Comparing the organic chemistry of the urban aerosols with the chemistry of the aerosols collected from smog-chamber irradiations of auto exhausts yielded no chemical basis by which to estimate the contribution of automobile exhaust to the total particulate burden. In general, the chemistry of the exhaust-related aerosols showed some significant similarities with the West Coast samples, marginal similarities with the New York samples, and no similarities with the Denver sample. In the end, it must be stressed that only a very limited number of aerosol samples have been analyzed for organic constituency. It is noteworthy however that good agreement between replicate analyses of a Pomona sample suggests that differences revealed by the organic analytical sequence are genuine. While it is felt that the samples and the analyses thereof are representative of the chemistry of urban atmospheric particulate, conclusion cannot be drawn about specific sources and precursors of the condensed organic matter.

RESULTS AND DISCUSSION

Light-Scattering Measurements

As discussed in the Background Section of this report, the quantity of haze was determined by an integrating nephelometer (with a heated inlet) which operates on the principle of light scattering. Throughout the course of this program, nephelometer measurements were made in New York City (Bronx, September, 1970; Welfare Island, August, 1972), Columbus (July, 1972), and Dayton (August, 1974), and the Blue Ridge Mountains (October, 1970). Data collected continuously at the Welfare Island and Dayton sites are of sufficient duration (nearly 1 month) to justify making generalizations about the average or typical patterns of haze in these localities.

The averages of the hourly averages of light scattering are plotted in Figure 3 for each hour of the day in Dayton and New York. The structure of the two diurnal light-scattering curves are quite similar. In both cities there is an increase in light scattering beginning near 1700 hours and continuing to a maximum in the early morning. In New York, and to a lesser extent in Dayton, there is a secondary maxima occurring during the morning rush hours. Surprisingly, there is a lack of a maxima during the noon or afternoon hours attributable to increased photochemical aerosol formation so commonly observed on the West Coast. Even on days where substantial photochemical smog was evident, and there were many (O_3 exceeded the 1-hour maximum standard of 0.08 ppm on 15 of 21 days in New York and 12 of 20 days in Dayton), afternoon increases in light scattering were only slightly pronounced. In the Blue Ridge Mountains where primary aerosol sources were negligible and light scattering sometimes approached background conditions (i.e., molecular scatter), an afternoon maximum in light scattering often occurred due to photochemical reactions involving terpenes and nitrogen oxides.

We conclude from these data that throughout an average diurnal period, light-scattering aerosols in midwestern and eastern cities originate primarily from primary emissions and/or dark (thermal) reactions of gaseous pollutants and secondarily from photochemical reactions. However, during the daytime when haze is visible, the two sources of aerosol are more nearly equal, and our conclusion regarding the principal source of these aerosols is made with caution, surmising that if it were not for the good ventilation of these cities in the afternoon, the photochemical aerosol contribution would be of greater significance.

Statistical Analysis of Light Scattering

The objective of this portion of the program was to develop an empirical model which would mathematically account for haze (light scattering) in terms of various air quality and meteorological data. It was desirable that the model be derived from hourly or shorter-time averages of the data; i.e., the model was designed to describe hour-by-hour changes in light scattering rather than long-term trends. A model developed in such a manner could perform a two-fold function:

(1) It could lead to the development of the capability for short-term forecasting of urban visibility and aerosol concentrations (much like a weather forecast).

(2) Since the basic time frame is short enough to reflect the impact of thermal and photochemical reactions, an empirical model of this type could yield information relevant to the understanding of these processes. Eventually,

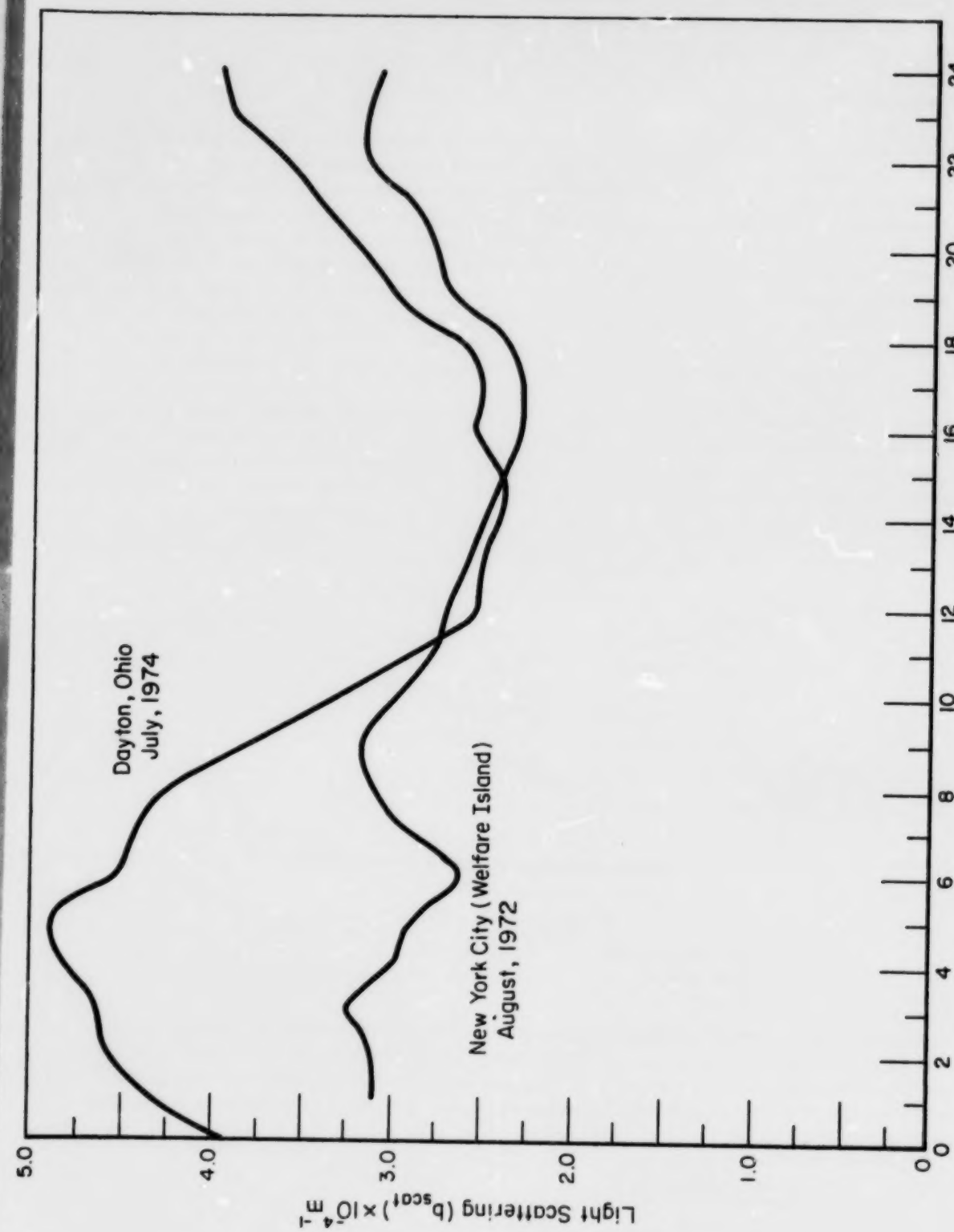


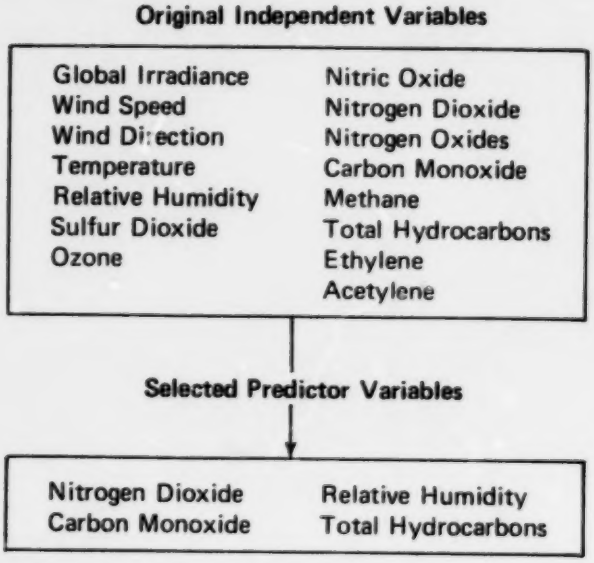
FIGURE 3. AVERAGE DIURNAL PATTERN OF LIGHT SCATTERING IN NEW YORK CITY AND DAYTON, OHIO

perhaps a blend of stochastic and theoretical chemical models could be developed to adequately explain the process of haze formation.

Preliminary Analyses

On the premise that the concentration of light-scattering aerosols at any given time might depend on the immediate history of the atmosphere, as well as some instantaneous factors, the approach selected for modeling light scattering was based on time-series analysis. Since the computer programs available for developing time-series models are limited to the use of three predictor variables, it was necessary to reduce the number of possible predictor variables to be included in our analysis. For this purpose, a multivariate empirical technique known as AID (short for Automatic Interaction Detector) was employed.(31) By means of an iterative binary splitting procedure, this method shows those variables and interactions among variables which are likely to best explain the variance of a criterion variable (in this case, light scattering). A more complete discussion of this procedure and a detailed account of the results is presented in Appendix A.

In the diagram below, the results of applying AID to the New York data are summarized. Of the 15 independent variables, NO_2, CO, relative humidity (RH), and total hydrocarbons (THC) were chosen as the variables to include in the time-series analysis. Since four variables were selected, it was decided to develop four time-series models for light scattering, one for each of the four possible combinations of three variables taken from this set of predictors.



It should be noted that time-lagged variables were not explicitly included in the AID analysis. However, since the variables were hourly averaged, there is some reflection of short-term time lag relationships in the data. Since time lags of even longer duration may be inherent in the photochemical processes of haze formation, it is possible that other important variables were overlooked by this screening technique.

Time-Series Analysis

Nine days of hourly averaged data (the four predictor variables NO_2, CO, RH, and THC, and light scattering) from August 22 to August 30, 1972, in New York were programmed for time-series analysis. Some preliminary analyses were performed on data as originally collected, i.e., at 10-minute intervals. However, for reasons discussed below, the remainder of the analyses were performed on hourly averages of 10-minute data.

The method used to develop the time-series models is that of Box and Jenkins.[32] The method, while theoretically based, places emphasis on pragmatism and parsimony, and hence is a practical method for time-series analysis of real data. Further discussion of the Box-Jenkins methods and detailed results are also included in Appendix A. A general discussion of the basic procedures follows.

Before developing the relationship between light scattering and the various predictor variables, it is first necessary to obtain univariate models for each of the variables to be included in the equation. These models relate the observed series to an uncorrelated noise series. The noise series may be interpreted as the effect of other variables on the time structure of the observed series. This noise series is then due to effects unexplainable by the given observed series, and it is the noise series that embodies the stochastic character of the observed series, i.e., the part due to chance. After the univariate models are developed, the multivariate relationship is studied. The multivariate relationship is incorporated in a time-series model called a transfer-function model. It is a dynamic model, and incorporates structural relationships between the dependent variable and predictor variables, as well as a noise series relating to those effects still unexplained by the relationship. The univariate models and transfer-function models are capable of fitting seasonal effects, trends, and even characteristics of time series which can change with time, e.g., stochastic trends. Diurnal variations were detected in the observed series for light scattering and the chosen predictor variables, and the subsequent model development incorporated seasonal (diurnal) factors to account for this variation.

Initially, AID analyses and time-series analyses were performed on the 10-minute data. However, difficulties were encountered in developing transfer-function models. Careful examination revealed that part of the problem was due to the fact that changes due to "noise" were on the same order as real changes and thoroughly intermixed, hence "masking" statistical relationships in the data. In an attempt to reduce this problem, hourly averages of data were used. This would tend to "smooth" variations due to noise and thus put in sharper focus basic changes occurring in the data. The AID analyses for the 10-minute data and the hourly averages were similar, but now meaningful relationships emerged in the transfer-function models. At this point, it was decided to carry out the time-series analyses solely on hourly averages.

A total of eight transfer-function models were developed. The four models mentioned previously were developed for each of two cases: (1) with seasonal differencing of the data, and (2) with no seasonal differencing of the data. There were two reasons for developing the second set of four models. First, it was noted that when seasonal differencing was applied (and this was indicated by preliminary Box-Jenkins analysis) the resulting models had a large degree of "overfitting". That is, much of the effect of seasonal differencing was nullified by parameters occurring in the consequent models. Hence, in interests of simplicity, it was decided to develop the nonseasonally differenced models. A second consideration was that the Denver data, obtained from EPA for purposes of attempting forecasts, covered only a short period of time, and diurnal differencing was not appropriate. It should be noted here that seasonal parameters remained incorporated in the nonseasonally differenced models.

The best of the eight models developed from the New York data, in terms of reducing the variance of the noise factor (i.e., the unexplained effects), was the diurnally differenced model involving the predictor variables NO_2, CO, and RH. The diurnally differenced models as a group were superior to the nondiurnally differenced models.

Table 1 presents a comparison of the different models according to the variance of the noise component. For a base comparison, the univariate light-scattering model and its variance are also included. Note that the NO_2, CO, and RH transfer-function model reduces the variance of the noise component by about 50 percent compared to the univariate light-scattering model.

It is important to consider that no conclusions about causality can be drawn from the structure of the time-series models developed, or from the variables included in the various models. These variables were chosen because, statistically, they form a set of predictors which could best reduce the unexplained effects. A variable such as CO, for example, is undoubtedly a "tracer" variable, that is, not one which is directly involved in the process of haze formation, but one which in some way summarizes information from other variables which are involved in this process. In the present analysis, therefore, the models developed are best suited for predictive purposes, and the tie-in between the empirical method of time-series analysis and theoretical cause-effect models is left to future research.

Forecasting

The transfer-function models developed form the basis for the forecasting procedure. Forecasts from a given point in time make use of past history of the light-scattering and predictor-variable series. The past history used for each variable depends on the structure of the particular forecast model. In general, a time period of up to several hours preceding the forecast is used, and if diurnal factors are included, appropriate information from preceding days are used. The forecast is comprised of two basic steps: first, relevant observed past values are put into the forecast model; second, the structure of the time-series model transforms this information into a forecast value for a given lead time.

Forecasts were of interest for two reasons: (1) various statistical tests performed on the forecasts could be used to attempt to validate models developed; (2) results of the forecasts could be studied to see how effective short-term forecasts for various lead times would be. Another consideration in the forecasting analysis was whether the models developed in New York City could be applied elsewhere. Data were obtained from Denver, Colorado, and Dayton, Ohio, for this purpose.

Figure 4 displays the results of the 1-hour-in-advance (lead-1) forecasts for light scattering over a 49-hour period from August 10 to August 14, 1972, in New York City. This period was not used in developing the forecast model. The transfer-function model in this case was the seasonally differenced model with predictor variables NO_2, CO, and RH.

The X-axis represents time in hours, the Y-axis light scattering in conventional b_{scat} units (10^{-4} m^{-1}). The solid line records the observed light scattering for this time period. The asterisks represent forecasts made 1 hour in advance, and the dotted lines designate the 50 percent confidence band about the forecasts, i.e., approximately 50 percent of the time the observed values should fall within the dotted lines. The figure was constructed in such a way as to provide immediate comparison between forecast values and the "eventual" observed values, and to yield visual evidence that the appropriate percentage of observed values fall within the confidence bands.

TABLE 1. COMPARISON OF UNIVARIATE AND TRANSFER FUNCTION MODELS OF LIGHT SCATTERING BY RESIDUAL VARIANCE

Model Type	Residual Variance in Light Scattering (x $10^{-4} m^{-1}$)
A. Models Without Seasonal Differencing	
Univariate model	0.300
Transfer function models with predictor variables	
I. NO_2, CO, RH	0.155
II. NO_2, CO, THC	0.169
III. NO_2, RH, THC	0.172
IV. CO, RH, THC	0.165
B. Models With Seasonal Differencing	
Univariate model	0.282
Transfer function models with predictor variables	
I. NO_2, CO, RH	0.146
II. NO_2, CO, THC	0.155
III. NO_2, RH, THC	0.182
IV. CO, RH, THC	0.157

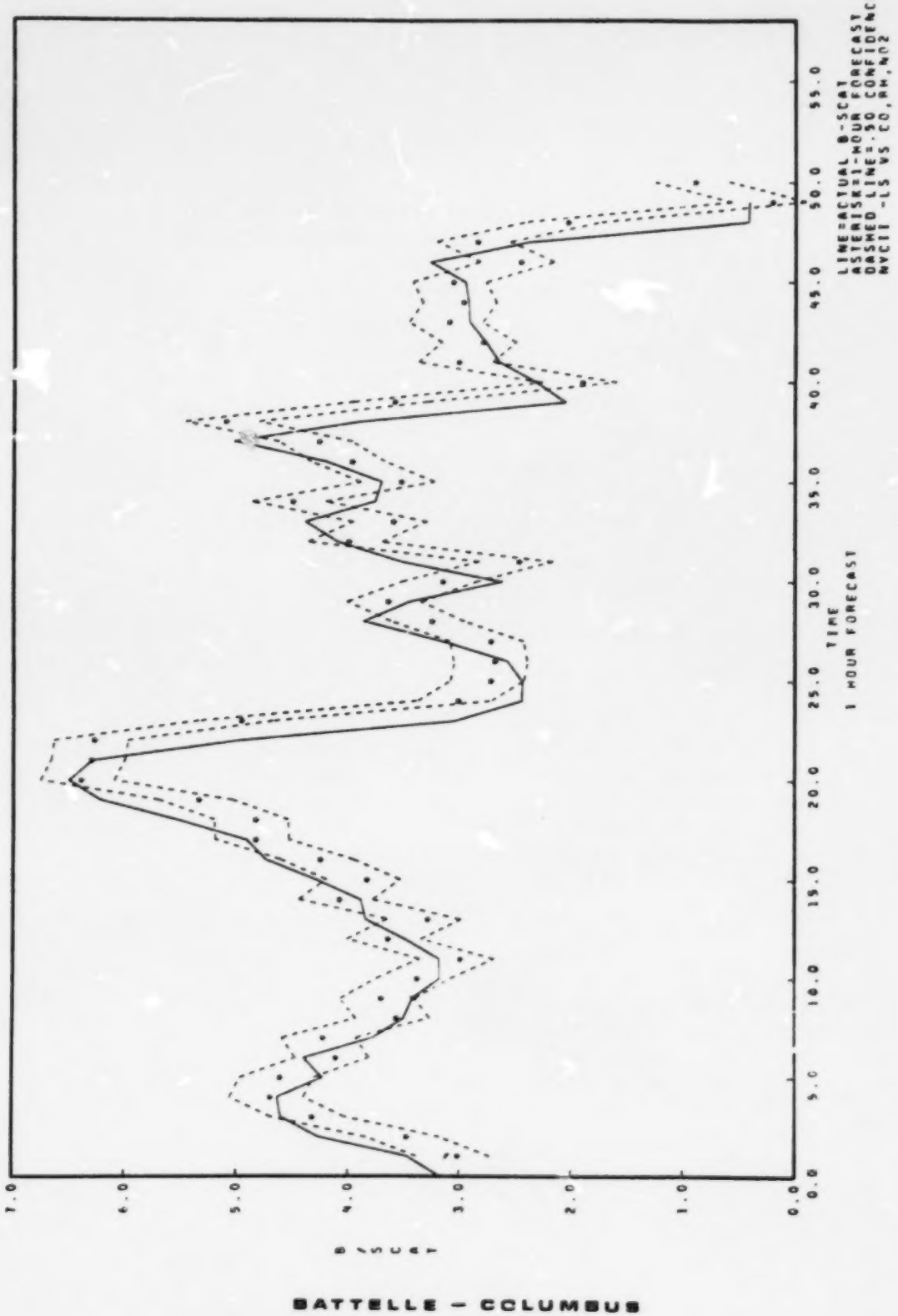

LINE=ACTUAL B-SCAT
ASTERISK=1-HOUR FORECAST
DASHED LINE=.50 CONFIDENCE BAND
NYCII -LS VS CO.,MH.,NO2
0.0
5.0
10.0
15.0
20.0
25.0
30.0
35.0
40.0
45.0
50.0
55.0
TIME
1 HOUR FORECAST
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
BATTELLE — COLUMBUS

FIGURE 4.

The width of the confidence band about the forecast in Figure 4 is ±0.33 b_{scat} units, and 24 out of 49 observed values fall within the confidence limits. Another statistical test shows that the 1-hour-lead forecasts are uncorrelated, a property they should possess. Hence, statistically and practically, the 1-hour lead forecasts are satisfactory.

Figures 5 through 7 show, respectively, the results of the 2-hour, 5-hour, and 8-hour lead forecasts using this model for predictions in New York. Table 2 summarizes the results of these forecasts (as well as those for other locations), including widths of confidence bands and percentage of observations falling within the 50 percent confidence bands. While perhaps still being acceptable, the longer lead forecasts may indicate some degradation in model effectiveness. Figure 8 is a composite of the lead-1 through lead-8 forecasts. The numbers 1 through 8 on the graph are located at the values of the 1-hour through 8-hour-in-advance forecasts, respectively. This figure shows the tendency of the forecasts to converge to the observed values as the lead time becomes shorter. The convergence illustrates the adaptive character of the Box-Jenkins forecasting model.

Figures 9 through 13 show the same information for forecasts made on data collected in Denver, Colorado, on November 15 and 16, 1973. The figures show about the same degree of success with the Denver data as with the New York data. However, because of the short span of consecutive data, and because the NO_2 data were incomplete, the forecast model used for Denver was nonseasonally differenced, and the predictor variables were CO, RH, and THC.

Finally, Figures 14 through 18 summarize the results of forecasts made on data collected in Dayton, Ohio, in July and August, 1974, where the seasonally differenced model was used with predictors NO_2, CO, and RH. Results are extremely good, perhaps indicating that a model developed from Dayton data would have a smaller noise variance than that developed from New York data.

Summary

The results of forecasting light scattering in three localities for lead times up to 8 hours show statistically acceptable results for the shorter lead times, with "reasonable" results for longer lead times. Further, the adaptive nature of the forecasting procedure continually improves forecast results as further information becomes available (and lead times become correspondingly smaller). The models used to generate the forecasts must be considered extremely simple and yet results are surprisingly effective. A possible explanation of why the forecasting method developed in New York seems to apply equally well in Dayton and Denver is that perhaps not enough sophistication has yet been built into the model, and that differences between individual cities are still embedded in the noise process. Further resolution into other predictor variables may make differences more pronounced.

One element of surprise in determining good predictor variables was that wind speed did not seem to be of much help in predicting light scattering. It was decided to investigate the data for a possible reason. Looking at wind speed in conjunction with wind direction showed more promise. Hence a method was devised to incorporate a wind-velocity vector into the AID analysis. Results, graphically displayed in Appendix A, show that wind velocity is an important predictor of light scattering. Unfortunately, it was not possible at this time to incorporate a vector-valued variable in the time-series analysis.

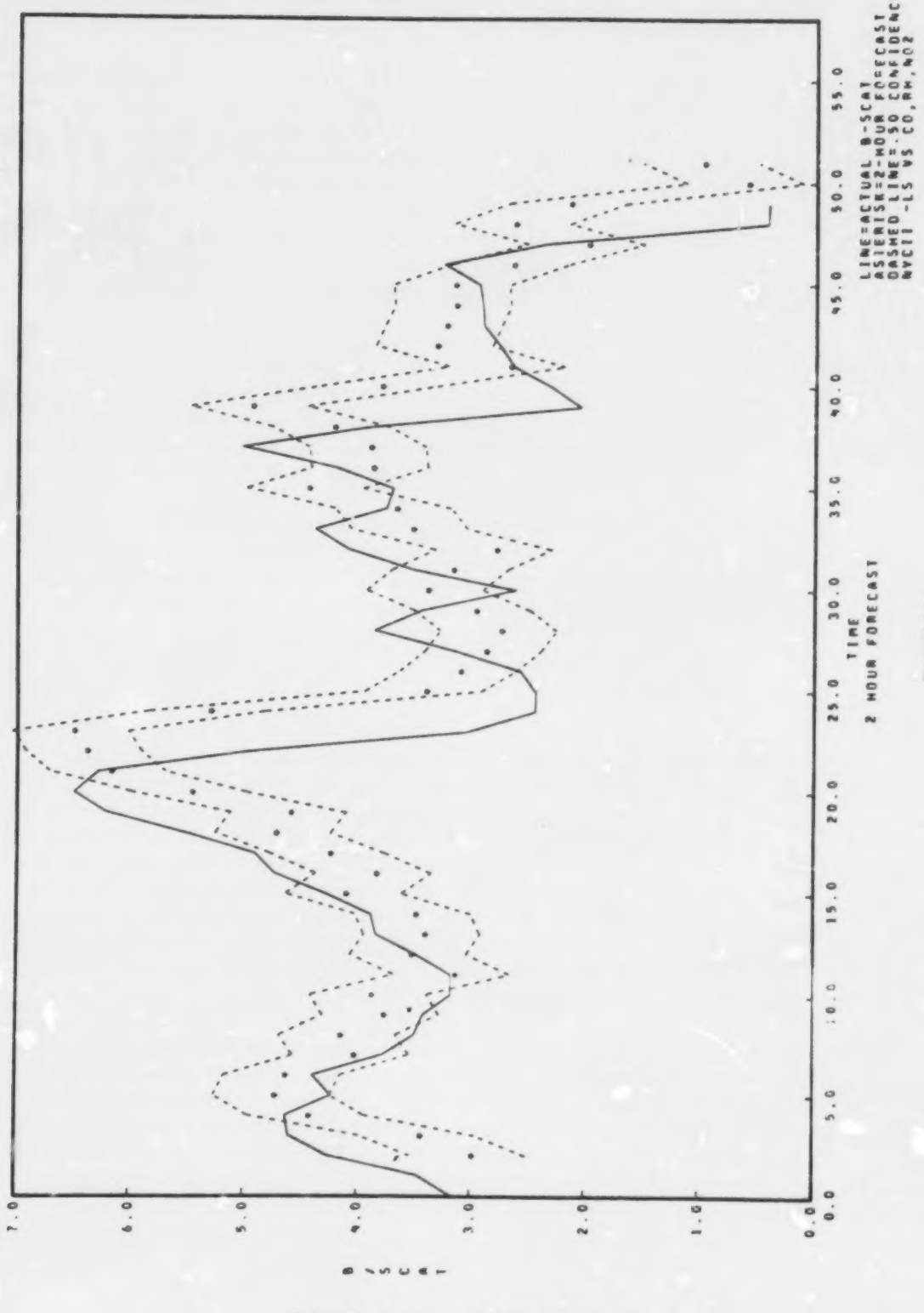

LINE=ACTUAL B-SCAT
ASTERISK=2-HOUR FORECAST
DASHED LINES=.50 CONFIDENCE BAND
NYCII -LS VS CO,RM,NO2
TIME
2 HOUR FORECAST
BATTELLE — COLUMBUS

FIGURE 5.

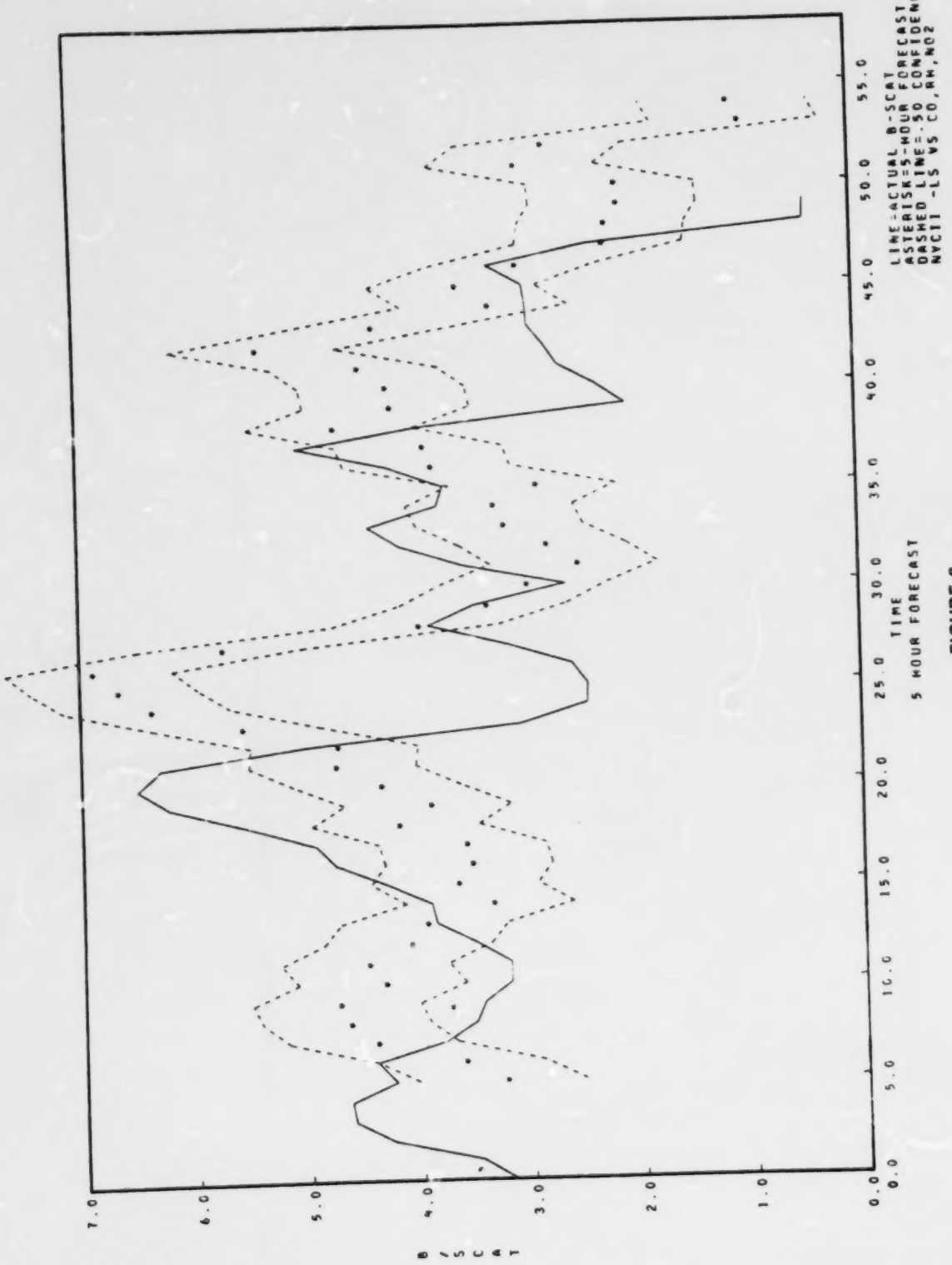

LINE=ACTUAL B-SCAT
ASTERISK=5-HOUR FORECAST
DASHED LINE=.50 CONFIDENCE BAND
NYCII -LS VS CO,RH,NO2
B SCAT
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
0.0
5.0
10.0
15.0
20.0
25.0
30.0
35.0
40.0
45.0
50.0
55.0
TIME
5 HOUR FORECAST

FIGURE 6.

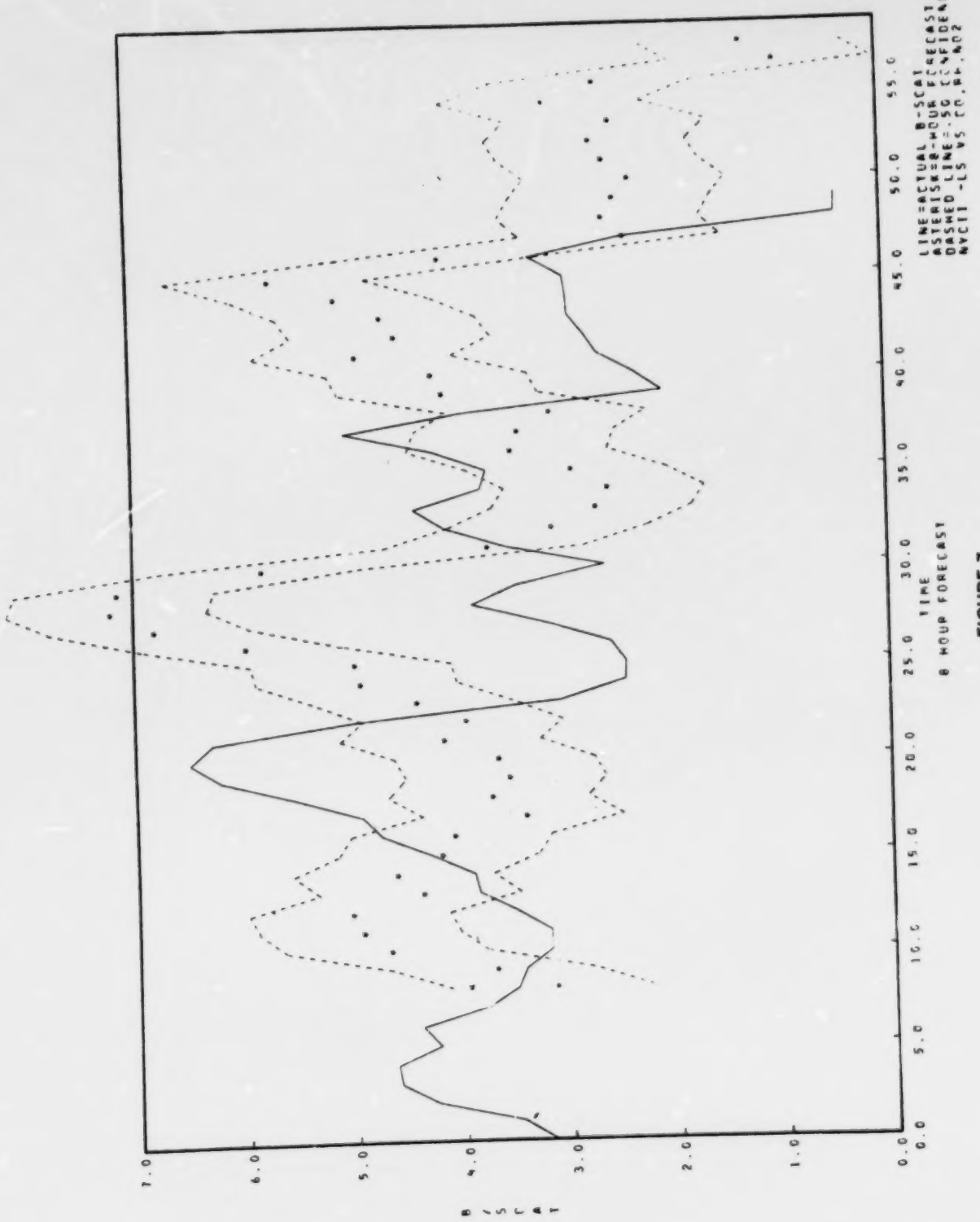

LINE=ACTUAL B-SCAT
ASTERISK=8-HOUR FORECAST
DASHED LINE=.50 CONFIDENCE BAND
NYC11 -LS VS CO.84-.N02
B-SCAT
TIME
8 HOUR FORECAST
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
5.0
10.0
15.0
20.0
25.0
30.0
35.0
40.0
45.0
50.0
55.0

FIGURE 7.

TABLE 2. FORECAST SUMMARY

Forecast Hours in Advance	.50 Confidence Band	No. of Observations in Forecast Band/Total
	NYC II	
1	± .33	24/49
2	± .51	22/48
3	± .60	
4	± .68	
5	± .75	16/45
6	± .81	
7	± .87	
8	± .93	12/42
	Denver	
1	± .33	21/38
2	± .52	20/37
3	± .62	
4	± .71	
5	± .785	14/34
6	± .855	
7	± .92	
8	± .98	9/31
	Dayton	
1	± .33	30/49
2	± .51	28/48
3	± .60	
4	± .68	
5	± .75	29/45
6	± .81	
7	± .87	
8	± .93	24/42

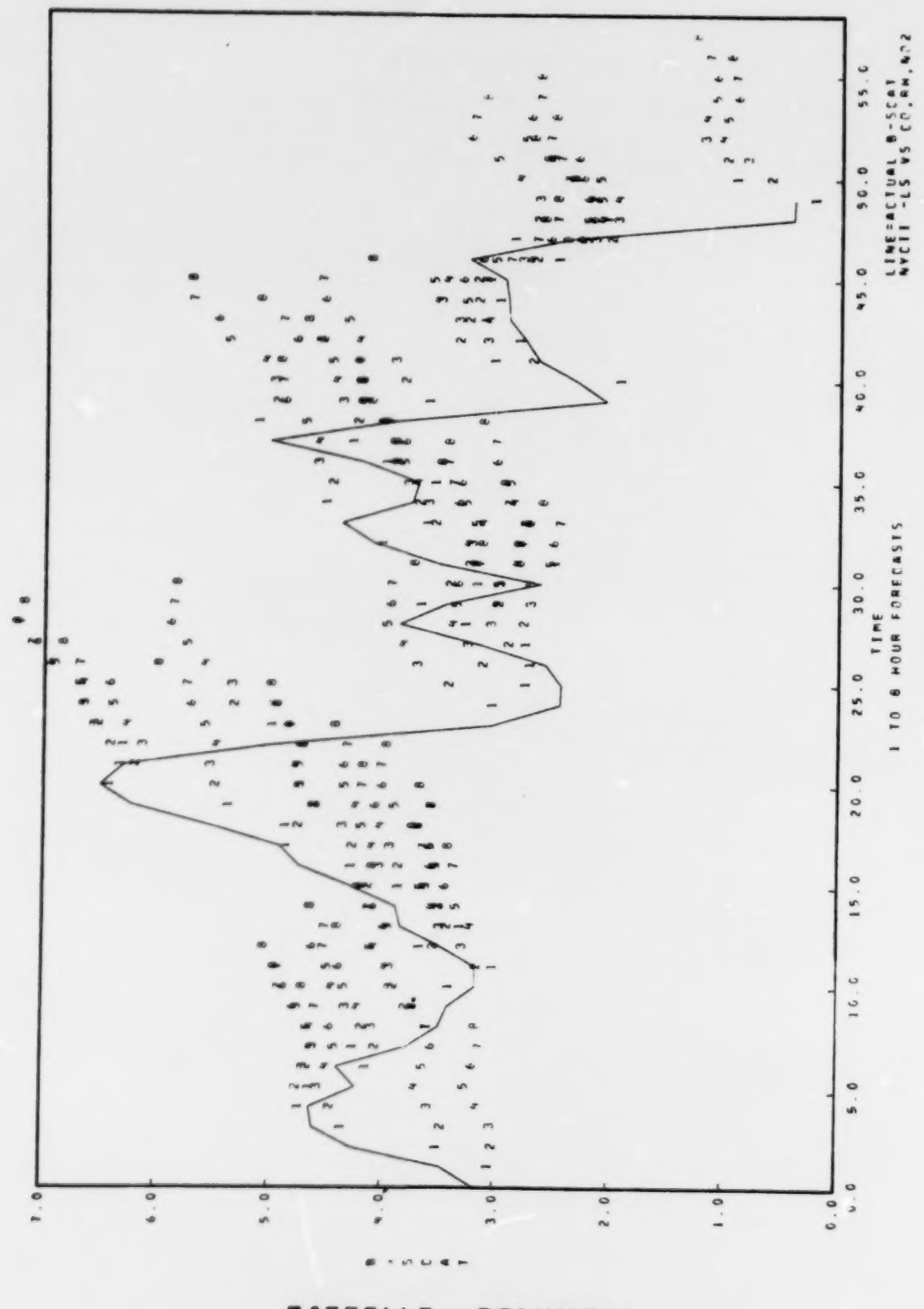

LINE=ACTUAL B-SCAT
NYCII -LS VS CO,RH,WS2
TIME
1 TO 6 HOUR FORECASTS
B-SCAT
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
5.0
10.0
15.0
20.0
25.0
30.0
35.0
40.0
45.0
50.0
55.0
BATTELLE — COLUMBUS

FIGURE 8.

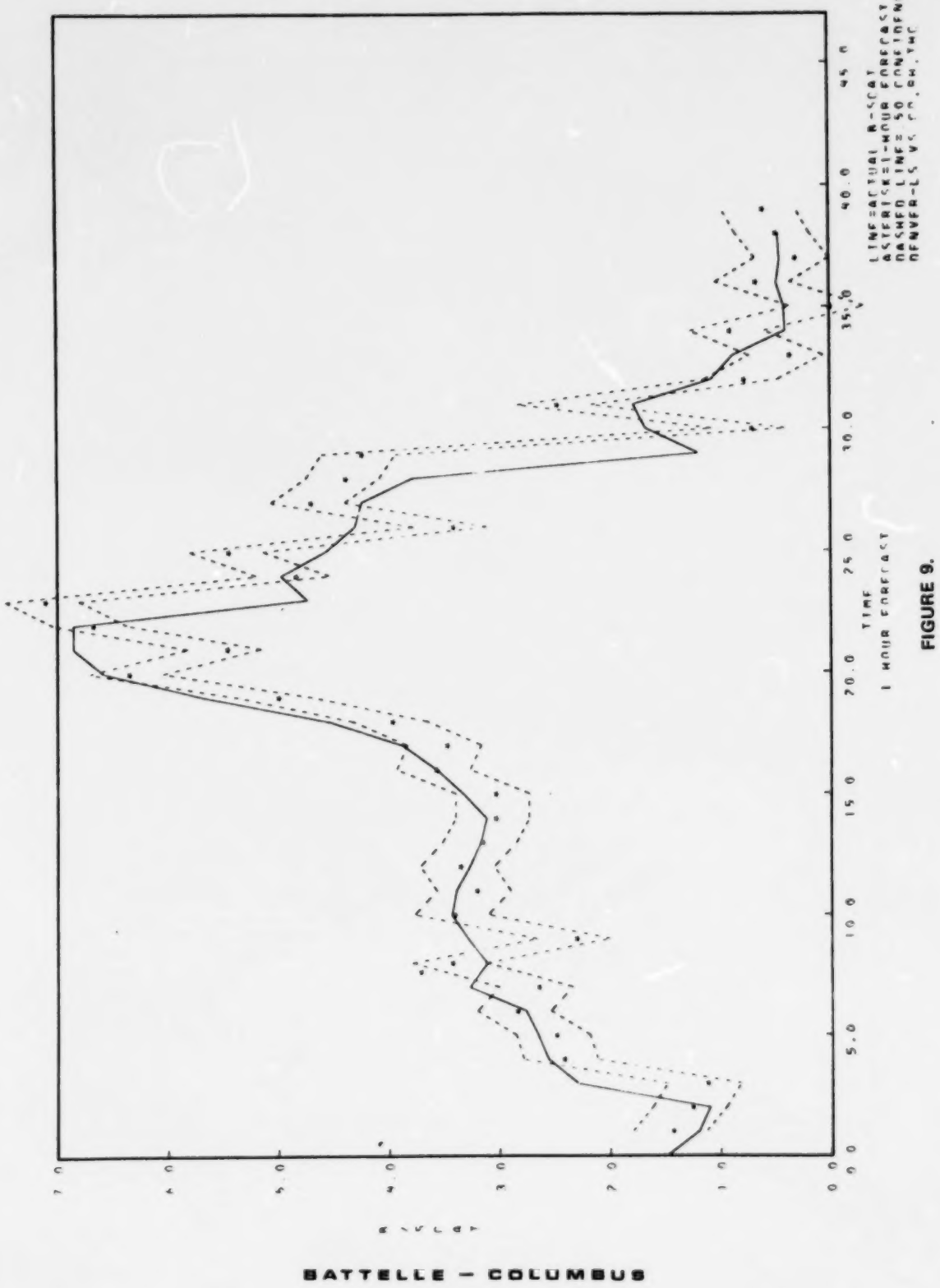

LINE=ACTUAL N-SCRAT
ASTERISKS=1-HOUR FORECAST
DASHED LINES=90 CONFIDENCE BAND
DENVER-LS VS PR.SH.THC
40.0
45.0
35.0
30.0
25.0
20.0
15.0
10.0
5.0
0.0
TIME
1 HOUR FORECAST

FIGURE 9.

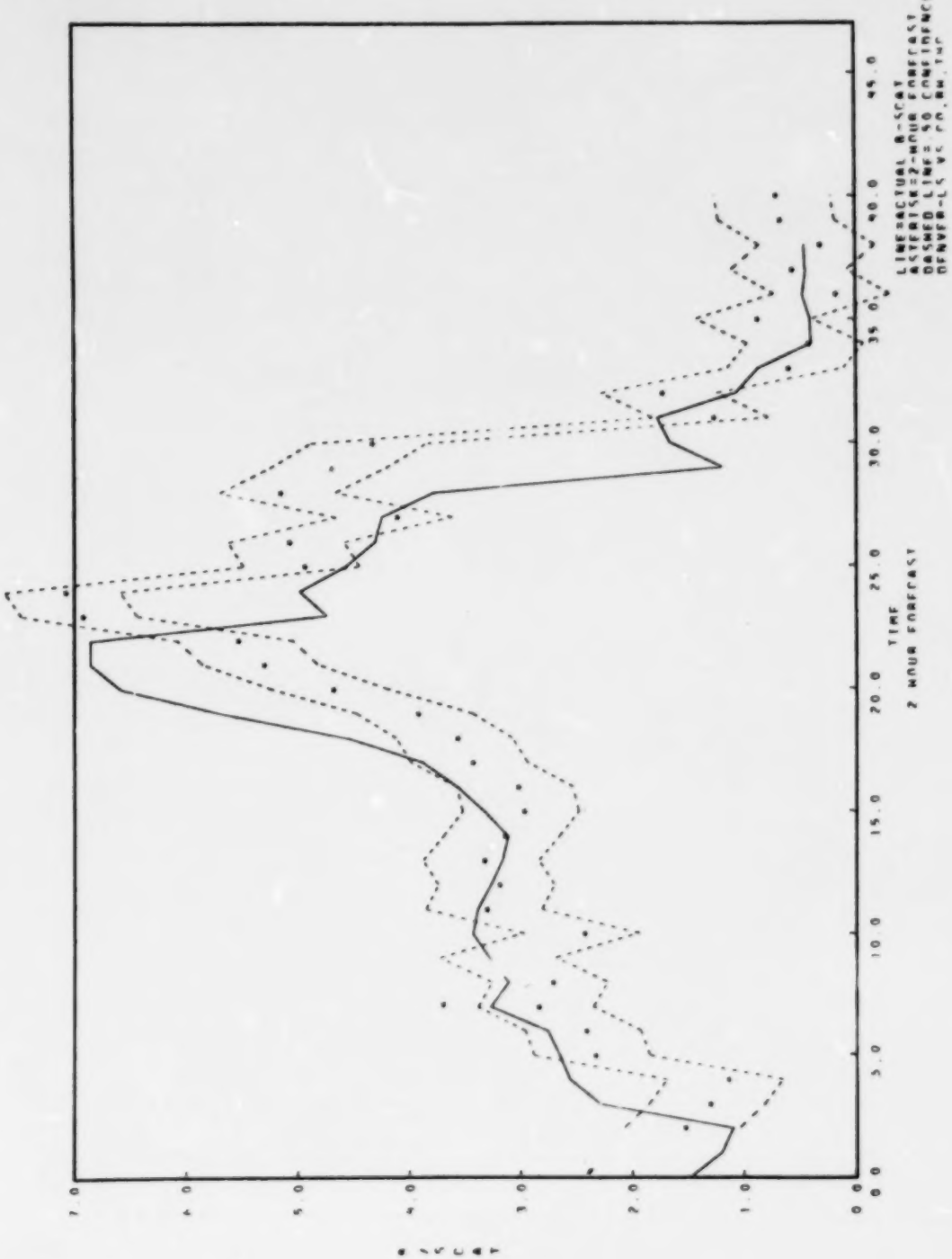

LINE=ACTUAL R-SCAT
ASTERISK=2-HOUR FORECAST
DASHED LINE= 90 CONFIDENCE BAND
0.0
5.0
10.0
15.0
20.0
25.0
30.0
35.0
40.0
45.0
TIME
2 HOUR FORECAST

FIGURE 10.

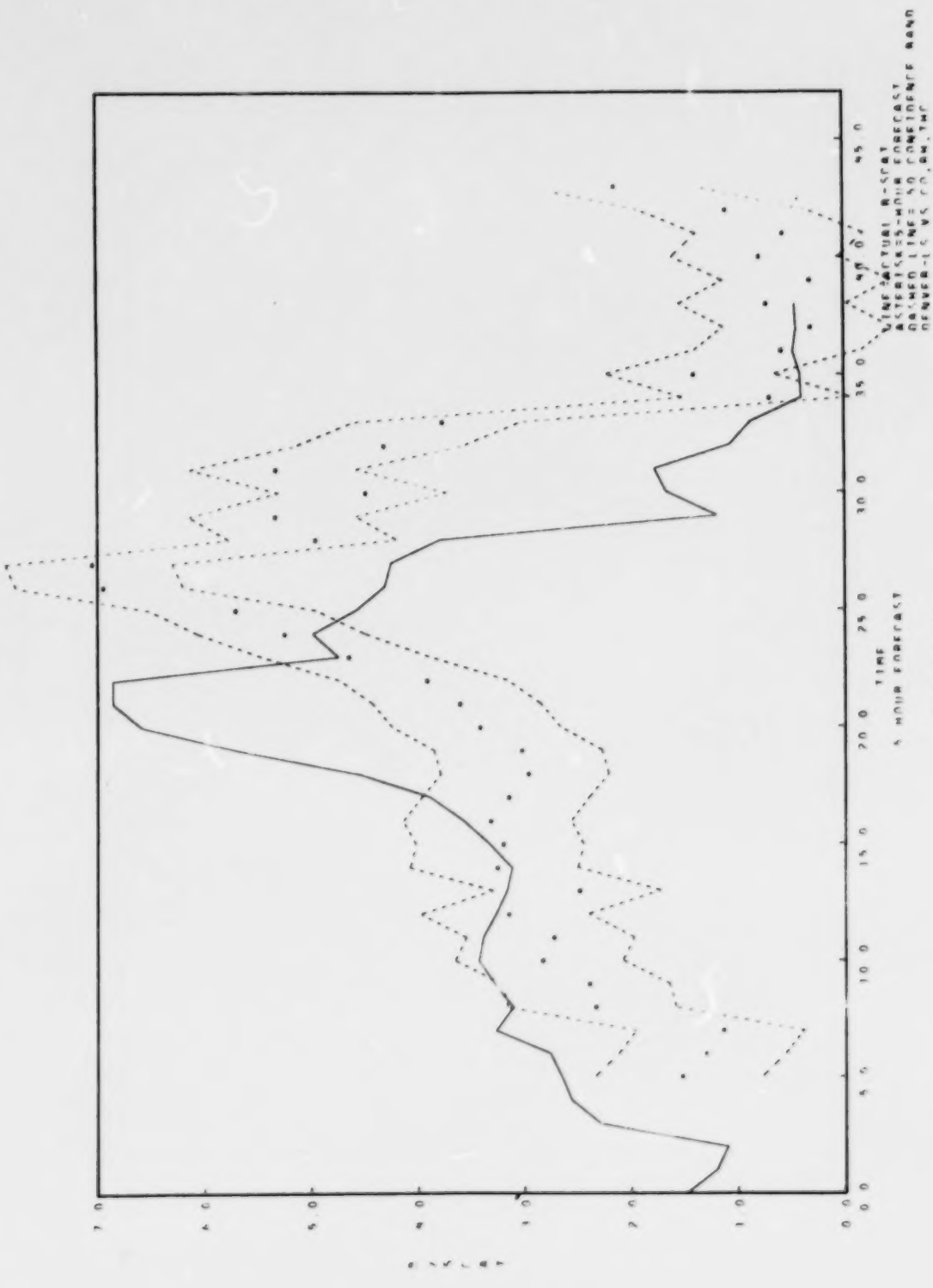


FIGURE 11.

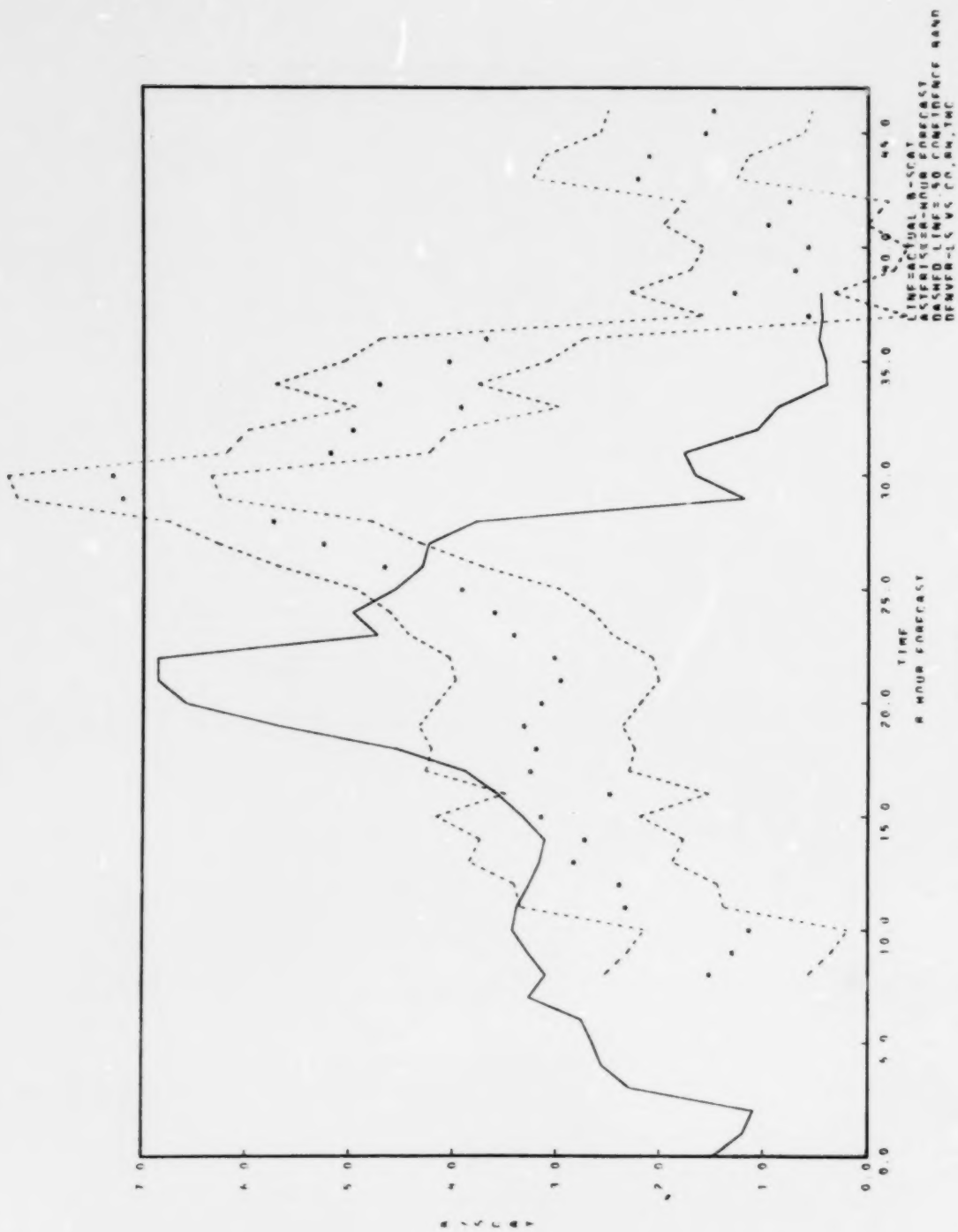

LINE=ACTUAL B-SCAT
ASTERISKS=4-HOUR FORECAST
DASHED LINES=90 CONFIDENCE BAND
DENVER-LV VS CP,84,THC
TIME
4 HOUR FORECAST

FIGURE 12.

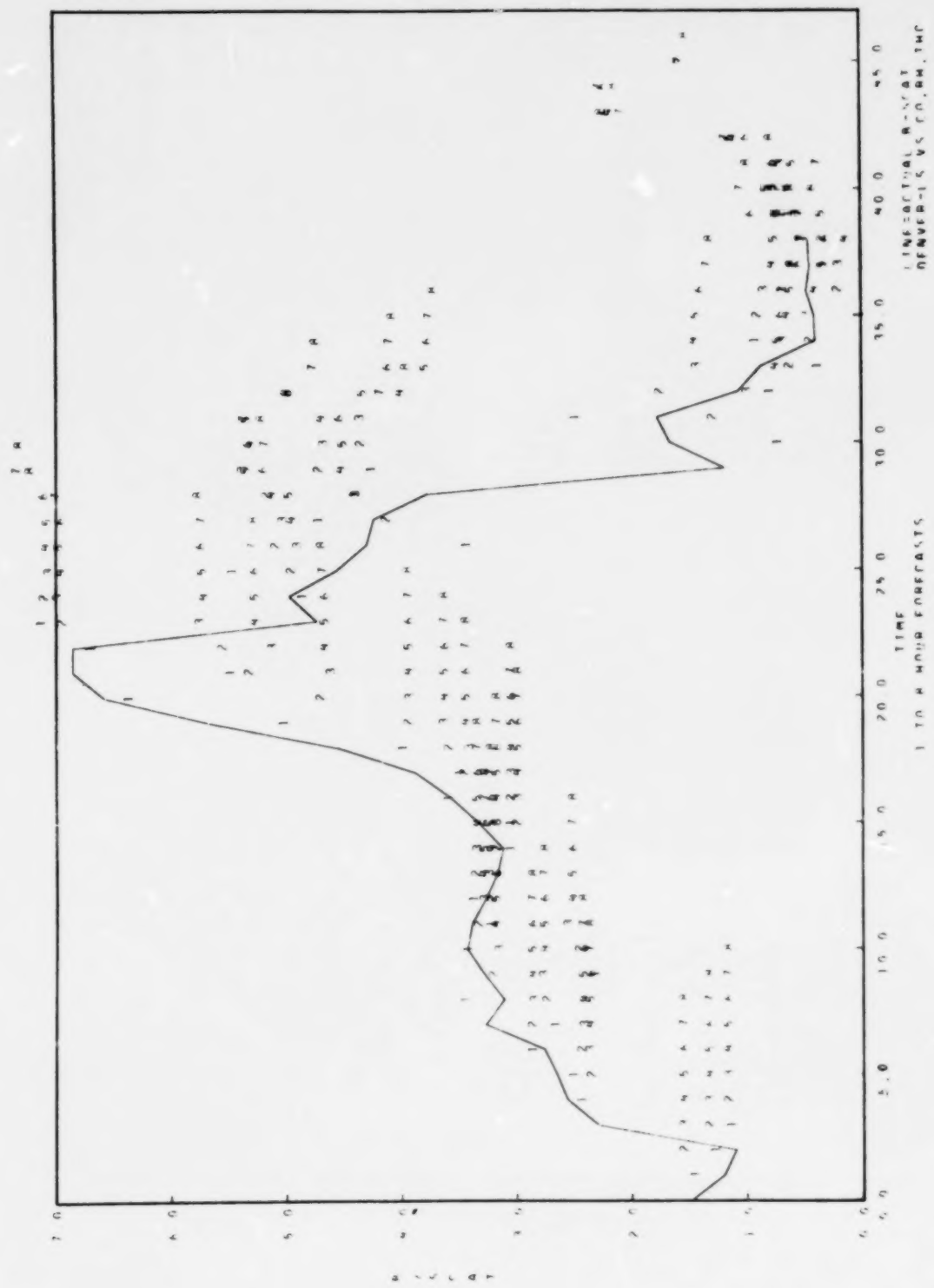


FIGURE 13.

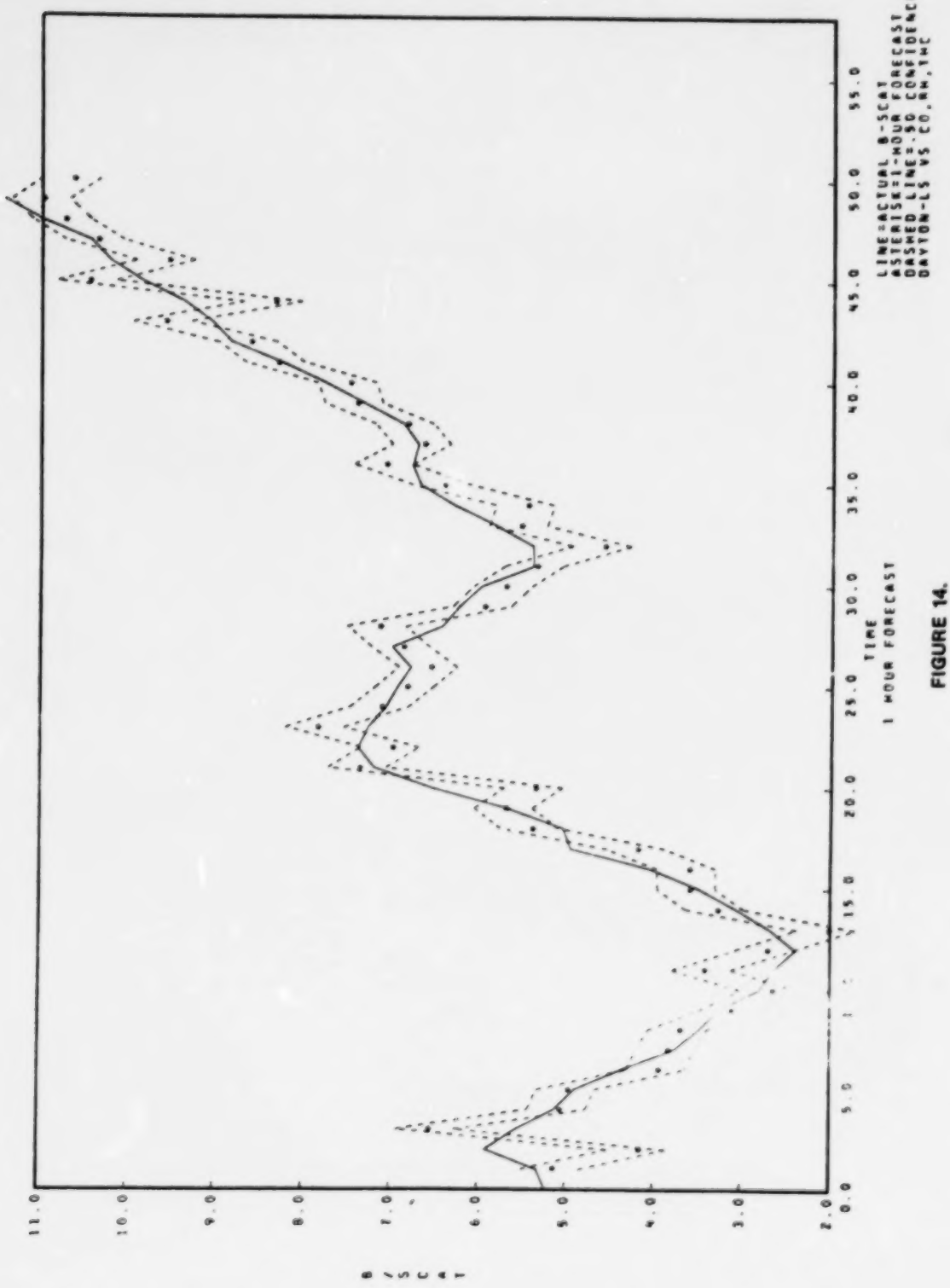

LINE=ACTUAL B-SCAT
ASTERISK=1-HOUR FORECAST
DASHED LINE=.90 CONFIDENCE BAND
DAYTON-LS VS CO.RM.THC
TIME
1 HOUR FORECAST
11.0
10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
0.0
5.0
15.0
20.0
25.0
30.0
35.0
40.0
45.0
50.0
55.0

FIGURE 14.

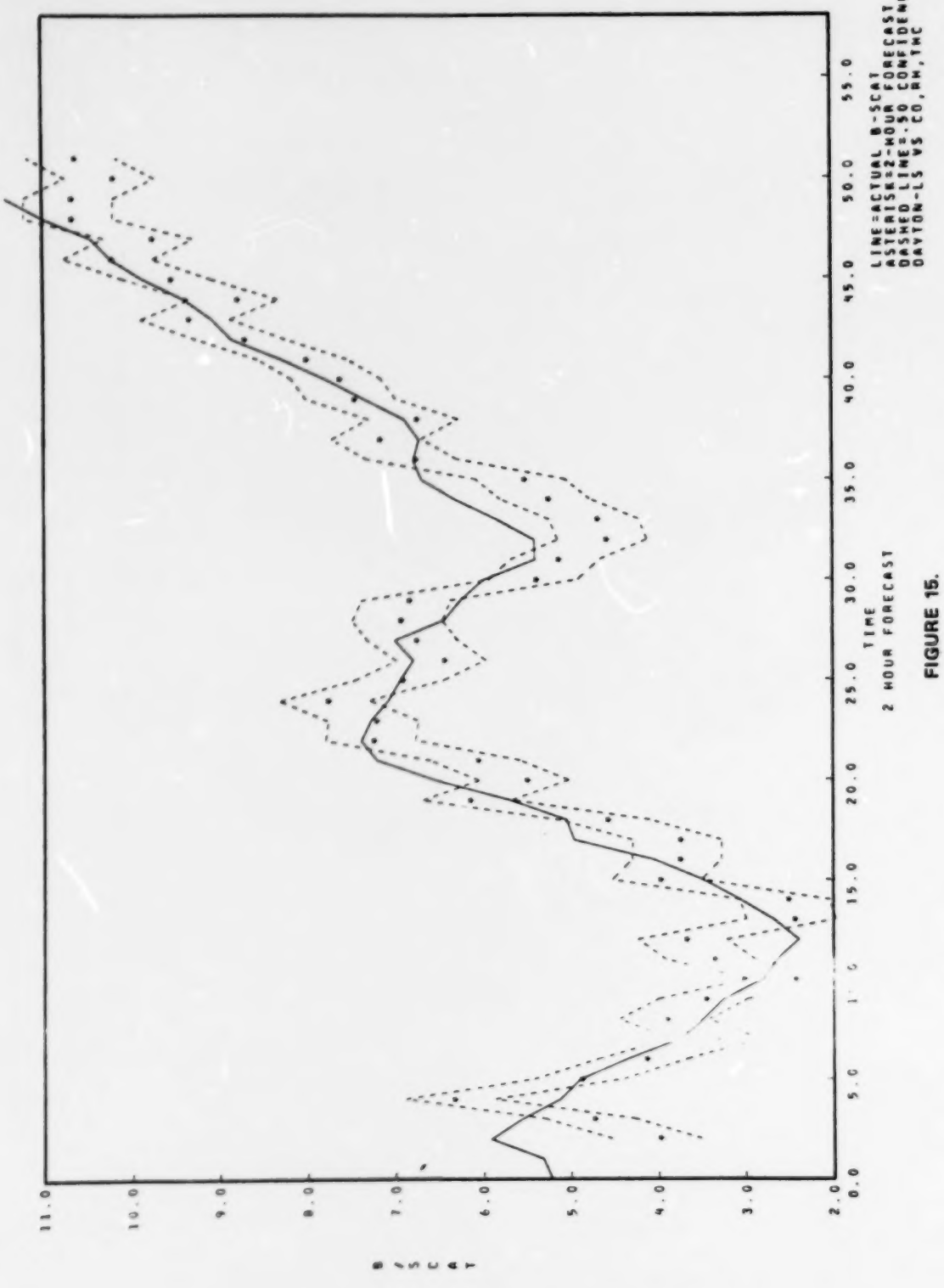

LINE=ACTUAL B-SCAT
ASTERISK=2-HOUR FORECAST
DASHED LINES=.90 CONFIDENCE BAND
DAYTON-LS VS CO,RH,THC
TIME
2 HOUR FORECAST
0.0
5.0
15.0
20.0
25.0
30.0
35.0
40.0
45.0
50.0
55.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
11.0

FIGURE 15.

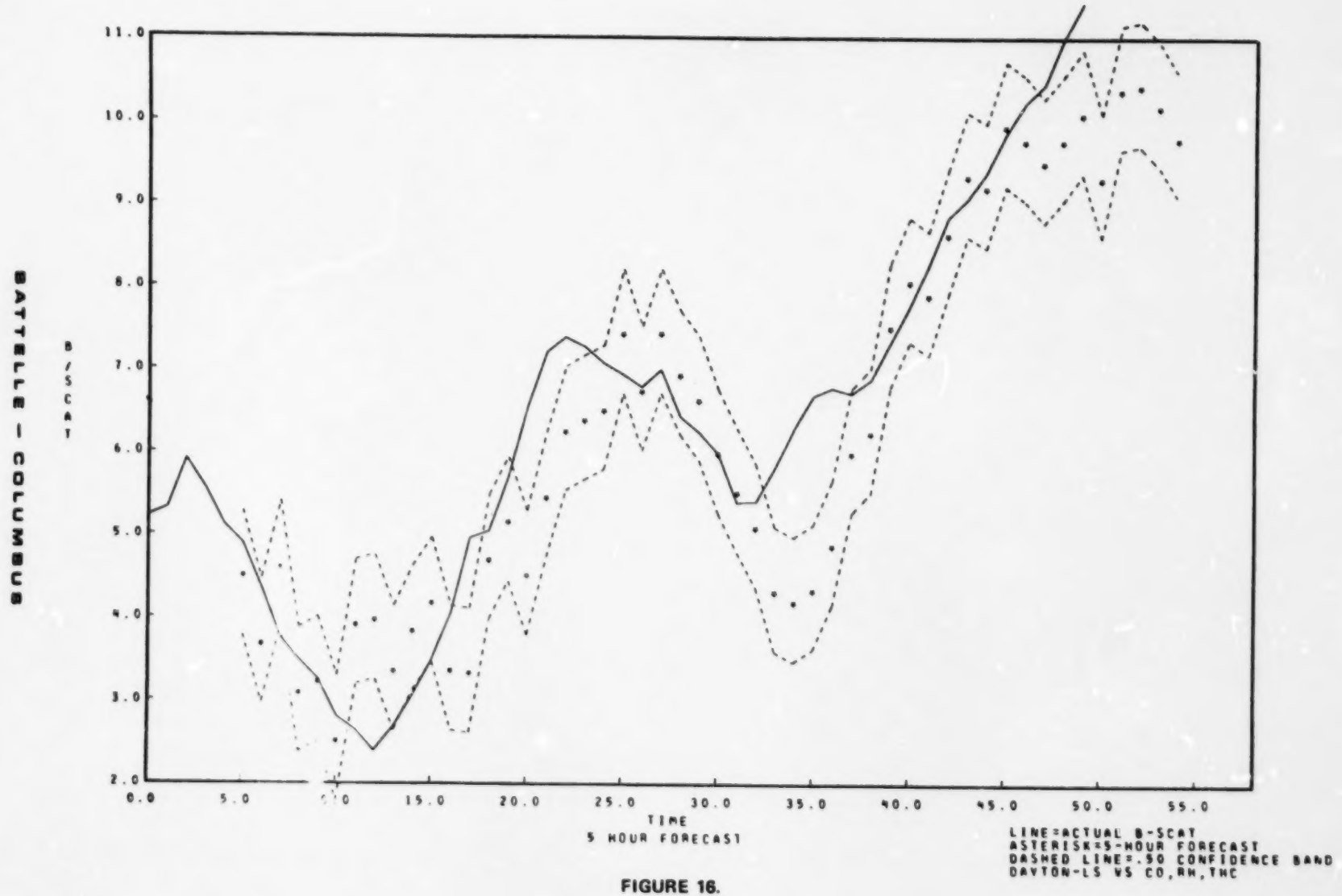

BATTELLE – COLUMBUS
B/SCAT
11.0
10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
0.0
5.0
15.0
20.0
25.0
30.0
35.0
40.0
45.0
50.0
55.0
TIME
5 HOUR FORECAST
LINE=ACTUAL B-SCAT
ASTERISK=5-HOUR FORECAST
DASHED LINE=.50 CONFIDENCE BAND
DAYTON-LS VS CO,RH,THC

FIGURE 16.

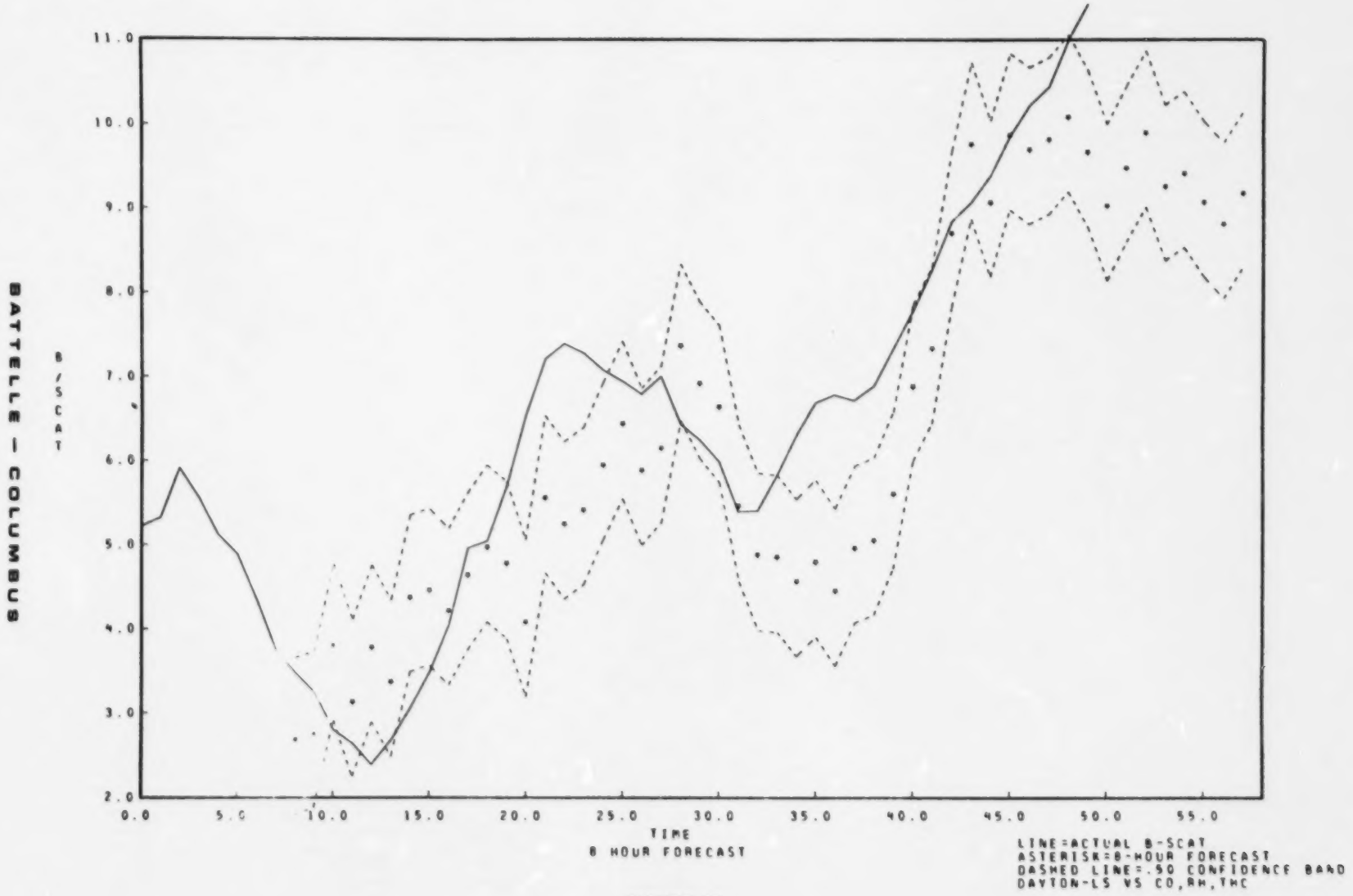

BATTELLE — COLUMBUS
B-SCAT
11.0
10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
0.0
5.0
10.0
15.0
20.0
25.0
30.0
35.0
40.0
45.0
50.0
55.0
TIME
8 HOUR FORECAST
LINE=ACTUAL B-SCAT
ASTERISK=8-HOUR FORECAST
DASHED LINE=.50 CONFIDENCE BAND
DAYTON-LS VS CO,RH,THC

FIGURE 17.

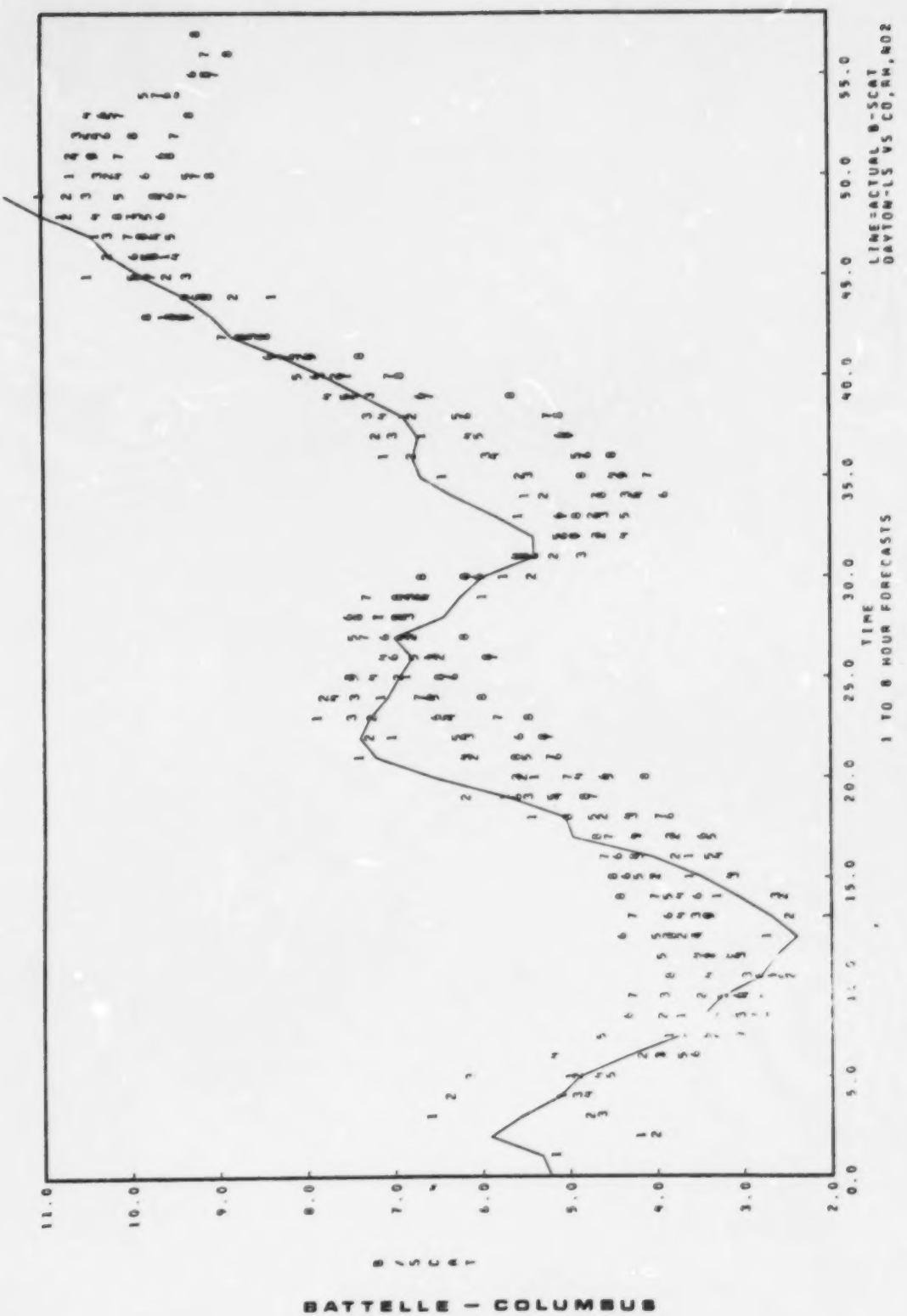

B-SCAT
11.0
10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
0.0
5.0
15.0
20.0
25.0
30.0
35.0
40.0
45.0
50.0
55.0
TIME
1 TO 6 HOUR FORECASTS
LINE=ACTUAL B-SCAT
DAYTON-LS VS CO,RH,402
BATTELLE — COLUMBUS

FIGURE 18.

Thus, although this model must be regarded as preliminary and incomplete, the results are very encouraging. It appears that with some additional expansion and refinement of the technique, it would be possible to make accurate predictions of the concentration of light-scattering aerosols in afternoons or evenings based on information gathered earlier in the day and from preceding days. One obvious improvement is to make provisions for including more predictor variables. Haze formation is a complicated process, and more than three variables are necessary to explain its evolution. A noticeable deficiency in the present models is the lack of more variables with a potential to describe the effect of time lags inherent in photochemical and perhaps thermal processes governing aerosol growth. Inclusion of such variables would be expected to enhance the performance of models in forecasting over longer lead times. Secondly, interactions among predictor variables may have profound effects on light scattering which do not show up when the variables are treated separately. A good example of this is the interaction between wind speed and wind direction. Taken individually, even though in succession, they do little to explain an effect on light scattering. Treated together, as a velocity vector, a powerful influence on light scattering is detected. Methods of identifying such interactions and incorporating them into the time-series analysis must be devised.

Chemical Composition of Light-Scattering Aerosols

Inorganic Composition

During the second year of this program, selected aerosol samples from Columbus, Ohio, New York City, and Pomona, California, were analyzed for their organic and inorganic compositions. Analyses were performed on diurnal collections of aerosols having diameters <2 μm. Analyses were also performed on composite samples of aerosols having diameters >2 μm. Detailed analytical results were presented in the Second Year Report.[2]

The average inorganic composition of the aerosols collected at the three sites is summarized in Table 3. As indicated in Table 3, there are vast differences in the composition of aerosols in the two size classifications implying that the sources of these aerosols are different and that there is little interaction (e.g., agglomeration) among aerosols in these size ranges; at least on a diurnal time scale. The chemical results tend to support the contention based on mobility measurements of aerosol size that aerosol mass in urban atmospheres is distributed bimodally, with a saddle point in the mass distribution near 2 μm diameter.[12] Indeed, it was on the presumed validity of that distribution data, as well as considerations of light-scattering properties, that particle separation was performed for diameters <2 μm.

According to the summary in Table 3, nearly all of the metallic compounds are associated with larger particles; of all the metals, only lead and zinc were found to have substantial fractions in the smaller aerosol size range. As a group, metallic elements constitute the major fractional part of the mass of larger aerosols, but <2 percent of the mass of the smaller aerosols. Soil-derived elements (Fe, Al, Si, and to a lesser extent Na, K, Ca and Mg) are found primarily among the larger aerosols. Elements of sea salt (Na, Cl, and to a lesser extent Mg) also predominate in the larger size range. Inorganic carbon compounds (carbonates) constitute the other major part of larger aerosols. Carbon in this case is assumed to be inorganic on the basis of the small amounts of hydrogen associated with the aerosols. The assumption is consistent with the results of Mueller, et al., who measured carbonate carbon in size fractionated aerosols from Pasadena.[33] They found that carbonate carbon was concentrated in size ranges >2 μm, and it constituted about 10 percent (12 $\mu g/m^3$) of the total particulate mass. If we assume that all of the carbon in the larger aerosols is inorganic, and all the carbon in the smaller size range is organic, then the inorganic carbon concentration in the New York area averages 9 $\mu g/m^3$.

TABLE 3. AVERAGE AEROSOL CONSTITUENTS AT THREE LOCALITIES [a]

	Columbus		New York		Pomona	
Aerosol diameter (μm):	<2	>2	<2	>2	<2	>2
Avg mass concentration (μg/m^3):	64	24	58	91	43	56
	Weight Percent According to Aerosol Size					
Constituent						
Metals	1	36	2	37	2	NC[b]
Carbon	17	17	26	16	25	"
Hydrogen	6	1	6	1	5	"
Halogen	0	1	1	6	2	"
Sulfate	19	2	22	4	14	"
Nitrate	1	2	4	3	24	"
Ammonium	6	0	7	0	9	"
Total	50	59	68	69	81	"

(a) The composition of aerosols <2 μm diameter are averages of samples from 5 days in Columbus, 11 days in New York, and 6 days in Pomona. Data for aerosols >2 μm are averages from 11 days in Columbus and 27 days in New York; no large aerosols were collected in Pomona.

(b) NC: large aerosols not collected for analysis.

Results of the same Pasadena study indicated that 90 percent of the organic carbon occurred in particles <1 μm in diameter, and that organic carbon comprised 37 to 54 percent of the particulate mass in that size range. By pyrolitic analyses, we found somewhat lower percentages of organic carbon in the smaller aerosols collected from Columbus, New York, and Pomona, but the very limited Pomona samplings are not likely representative of the Los Angeles area. As discussed in the next section of this report, organic extractions by methylene chloride and dioxane indicated 20-30 weight percent organics in the light-scattering aerosols around Columbus and New York.

Nitrate was found in both aerosol size ranges in contrast to a report by Lee and Patterson[18] where nitrate predominated in the submicron size range. Judging from recent evidence that gaseous nitric acid can be captured as particulate on ordinary glass fiber filters[34], it is suspected that discrepancies regarding the distribution and/or concentration of nitrate in aerosols may be related to nitric acid collection. High purity quartz filters (to our knowledge used here for the first time in atmospheric sampling), do not seem to absorb nitric acid, and the nitrate concentrations we report may therefore be somewhat less than those reported previously.

Ammonium was found to be exclusively associated with the smaller size fraction of aerosols which is consistent with most reports on the ammonium distribution. The sum of nitrogen from nitrate and ammonium accounts for about 75 percent of the total-aerosol nitrogen. The question is often raised whether ammonium salts are formed on filter substrates from reactions of gaseous ammonia with acids. This is an important consideration, and our results do not provide answers to that type question. Although we are reasonably certain that nitric acid did not "fix" ammonia on the quartz substrate, it is possible that sulfuric acid did.

Along with organic matter, sulfate compounds constituted a major fraction of the mass of the smaller aerosols. The exclusiveness of sulfate among small particles does not imply a dominant source or mechanism of sulfate formation as numerous processes for both heterogeneous and homogeneous conversion of SO_2 to sulfate have been postulated which result in submicron aerosols. Sulfate sulfur accounted for only about one-third of the total-aerosol sulfur, but no determinations were made for sulfite. Novakov et al. reported that sulfite sulfur predominated over sulfate sulfur in the <2 μm aerosol-size range.[24] The average concentration of sulfate in light-scattering aerosols in New York of 12 $\mu g/m^3$ is near the overall 5-year (1964-1968) average of 11.4 $\mu g/m^3$ of total sulfate for eastern urban sites.[35]

The halogen contribution to the mass of small aerosols was <2 percent. In New York, chloride was the predominant halogen – and its concentration is attributed primarily to sea salt.

Regression analyses were performed to determine statistical relationships which might exist between the inorganic composition of aerosols and the air quality and meteorological data associated with the respective periods of aerosol collections. Eleven days of sampling in New York was used for the analysis, and the air quality data were averaged on a 24-hour basis for comparison with the diurnal aerosol samples. The results are summarized in Table 4.

Before discussing the regressions relating aerosol composition to air quality data it is important to note that, in several instances, the regression variable selected as most significant on the basis of the highest correlation coefficient was only slightly better than other highly correlated independent variables. Correlations among the independent variables are indicated in Table 5; coefficients >0.75 appearing in bold face. Correlations among the dependent variables, i.e., the compositional components of the aerosols, were quite poor with the exception of the correlation between sulfate and ammonium concentrations which was $R = 0.94$.

TABLE 4. REGRESSION ANALYSES ON INORGANIC AEROSOL COMPOSITION AND AVERAGE AIR QUALITY CONDITIONS

Type of regression:	Linear Mass Concentration			Linear Percent Concentration			Logarithmic Mass Concentration		
Variable dimension: Dependent Variable	Variable[a]	R	F	Variable	R	F	Variable	R	F
Pb	NO_x	0.87	28.9	NO_x	0.90	32.9	NO_x	0.88	32.3
Br	NO	0.79	14.8	O_3	-0.75	11.4	O_3	-.77	13.2
Cl	C_2H_2	0.59	4.9	temp[b]	-0.51	3.2	C_2H_2	.59	4.9
S	b_{scat}	0.58	4.7	WD	0.42	1.9	b_{scat}	0.56	4.2
SO_4	b_{scat}	0.95	75.1	b_{scat}	0.78	13.9	b_{scat}	0.92	46.4
NO_3	C_2H_2	0.57	5.0	temp	-0.49	2.9	C_2H_2	-0.66	6.9
NH_4	b_{scat}	0.98	76.8	b_{scat}	0.83	20.2	b_{scat}	0.91	42.3
C	WS	-0.79	15.0	THC	-.50	3.0	WS	-0.81	17.5

(a) In several instances, the "best" independent variable had a correlation coefficient only slightly higher than that of another independent variable. Correlations between the independent variables are listed in Table 5.

(b) Underlined variables do not meet the 95 percent confidence test.

TABLE 5. CORRELATION COEFFICIENTS AMONG THE INDEPENDENT VARIABLES SUBSEQUENTLY REGRESSED AGAINST AEROSOL COMPOSITION

	SO_2	O_3	NO	NO_2	NO_x	CO	CH_4	THC	C_2H_2	NMHC	BSCAT	WS	TEMP	RH	IRRAD
SO_2	1.00														
O_3	-.46	1.00													
NO	.41	-.89	1.00												
NO_2	-.73	.23	-.02	1.00											
NO_x	-.31	-.39	.61	.74	1.00										
CO	-.55	.08	.17	.89	.80	1.00									
CH_4	-.55	.27	-.09	.74	.48	.88	1.00								
THC	-.60	-.03	.16	.85	.74	.68	.43	1.00							
C_2H_2	-.43	-.27	.23	.62	.60	.68	.61	.48	1.00						
NMHC	-.26	-.27	.29	.41	.50	.17	-.15	.80	.99	1.00					
BSCAT	-.31	.22	.05	.79	.67	.83	.65	.63	.39	.27	1.00				
WS	.25	-.01	-.16	.35	-.60	-.60	-.38	-.57	-.47	-.36	-.70	1.00			
TEMP	-.25	.56	-.23	.49	.25	.38	.18	.32	-.06	.17	.64	-.32	1.00		
RH	-.41	.20	.05	.73	.56	.75	.71	.62	.41	.24	.78	-.33	.49	1.00	
IRRAD	.37	-.01	-.2	-.73	-.74	-.90	-.79	-.60	-.59	-.18	-.82	.75	.59	-.69	1.00

BATTELLE - COLUMBUS

Referring to Table 4, three types of regressions were derived for each of the dependent variables appearing in the left-hand column. In the first case (column 2), a linear model was fitted using mass concentration units for the dependent variables. In the second case (column 3), the aerosol composition data are expressed as the percent of the total aerosol mass concentration, and in the final case, concentration units were used again but with an exponential regression model. In Table 4, R denotes the correlation coefficient and the sign of the relationship, and F denotes the degree of significance of the correlations; the significance increases with increasing F values. An F value of about 4 corresponds to a 95 percent confidence interval, and a value of 13.5 corresponds to a 99 percent confidence band.

The results obtained where linear and logarithmic equations were regressed on the same data are remarkably similar. In all but one case where the correlation coefficients exceed 0.75, the independent variables having the highest correlations were the same for the linear and logarithmic models, and the correlation coefficients were also quite similar for the two models. This implies, of course, that the correlations tend to be linear. In the one case (Br) where the most highly correlated variable was different for the two models, a substantial correlation exists between the two highest correlated predictor variables (R = -0.83 for O_3 versus NO).

Linear regression results where the compositional data are expressed on a percentage basis are not much different from those where absolute concentrations were used. Overall, the correlations with percent composition are lower, and in four cases (Cl, total S, NO_3, and total C) the regression equations are not significant at the 95 percent level.

On the basis of these analyses, there is definitely a statistically significant correlation between daily averaged Pb concentrations and the 24-hour average concentration of NO_x. One might expect a priori that the concentration of most any of these aerosol substances would correlate best with light scattering (b_{scat}) since aerosol mass loading (<2 μm) increases with increasing light scattering (R = 0.94 for b_{scat} versus mass <2 μm). This result occurs in the regressions on S, SO_4, and NH_4. The fact that Pb shows a better correlation with NO_x under these circumstances, further supports an apparently strong correlation between NO_x concentration and the amount of Pb in aerosols.

As noted, there is substantial correlation between SO_4 and NH_4 concentrations (total S to a much lesser extent) and average light scattering. Significant positive correlations between b_{scat} and the percent of SO_4 and NH_4 indicate that the percentage of $(NH_4)_2SO_4$* in the particulate tends to increase on days of highest light scattering.

Correlations of air quality data with the concentration of particulate Cl, total S, and NO_3 are quite low and therefore of little meaning. The particulate carbon concentration correlates fairly well with negative wind speed; again, an expected first-order correlation which offers little information on the possible relationships with sources.

To search further for statistical relationships on aerosol composition, stepwise multiple regression equations were derived. Statistically acceptable regression equations containing two or more independent variables were possible for Pb, total S, SO_4, NO_3, and NH_4. The results are summarized below. The correlation value (ΣR) there is the cumulative coefficient (required to increase with each succeeding step). The F values represent the significance of adding the final term; the significance of the regression equation is always much larger. A negative sign indicates negative correlation with the specified variable.

*The near 2/1 stoichiometric ratio of NH_4 to SO_4, and the high correlation between their concentrations lead us to assume that nearly all of the NH_4 and SO_4 form the $(NH_4)_2SO_4$ compound, in agreement with other reported work where the structure was confirmed by X-ray analyses.

Dependent Variable	Second Step			Third Step		
	Variable	ΣR	F	Variable	ΣR	F
Pb	-WS	0.93	7.25	NO_2	0.96	5.0
S	-RH	0.78	5.8			
SO_4	-NMHC	0.97	6.3			
NO_3	-NO	0.76	4.2			
NH_4	-NO	0.98	24.0	-RH	0.99	11.3

Including wind speed and NO_2 in the correlation on Pb improved the overall regression coefficient from an initial value of 0.87 to 0.96. The negative correlations appearing in the second regression steps on S, SO_4, NO_3, and NH_4 are obviously of no positive benefit in delineating any aerometric effects on aerosol composition. The negative correlation between NH_4 concentrations and RH tends to challenge the theory that $(NH_4)_2SO_4$ formation occurs predominantly in water droplets at high humidities. The concentration of carbon, for which wind speed correlated best in the first regression step, showed no second-order correlation that was significant.

Thus, with the exception of particulate Pb which correlated consistently with nitrogen oxide concentrations, no outstanding statistical evidence was uncovered which would allow us to draw conclusions regarding variations of the inorganic composition of aerosol. For the most part, short-term (i.e., daily) changes in inorganic composition are not great, and the rather subtle relationships which may exist remain hidden by the complexities of the atmospheric conditions.

Organic Composition

Beginning the second year of this program, methods in organic analyses were developed and/or adapted to provide information concerning the broad chemical classes and functional groups present in the organic fraction of light-scattering aerosols. Two major factors determined the nature of the analyses undertaken:

(1) The limited quantities of sample available for analysis

(2) The desirability of obtaining numerical (i.e., semiquantitative) data.

Because it is ultimately desirable to relate the organic composition of aerosols to diurnal trends in atmospheric conditions, samples of aerosols available for analyses were limited to an average mass of about 14 mg, i.e., the average mass of organic matter extracted from the <2 μm aerosol fractions collected from three sampling systems operated continuously at a flow of 20 cfm.

In the selection and development of procedures to be included in the organic analytical scheme, particular attention was given to the time and cost of each operation. In view of an ultimate goal of obtaining data on a statistically significant number of samples, the cost per sample could not be excessive. Nevertheless, while not prohibitive in cost, the procedures do require a moderate degree of analytical sophistication (refer Appendix B). During the second and third years of the program, the analytical scheme was applied to ten samples of atmospheric particulate and two samples of particulate corresponding specifically to automotive sources. The samples, listed below, have been numbered to permit convenient reference in the discussion to follow.

Sample Reference No.	Site/Source	Date
1	Columbus, Ohio	July 21, 1972
2	Columbus, Ohio	July 26, 1972
3	New York City	August 23, 1972
4	New York City	August 11, 1972
5	Pomona, California	November 18, 1972
6	Pomona, California	November 10-13, 1972
7	Pomona, California	October 14-18, 1972
7A	Replicate A	
7B	Replicate B	
8	Denver, Colorado	November 16-17, 1973
9	Rubidoux, California	September 21, 1973
10	West Covina, California	September 21, 1973
11	Primary Automotive Exhaust	
12	Irridiated Automotive Exhaust	

Where available, atmospheric data corresponding to the sample collection periods are summarized in Table 6. Sample Nos. 1 and 2 from Columbus, 3 and 4 from New York, and 5 and 6 from Pomona were collected on days of substantially different air quality, based primarily on the range of light-scattering values encountered during the sampling periods at each site. Columbus sample No. 1 was collected during one of the most severe smog episodes in the Ohio area. New York sample No. 3 was selected as a typically hazy day in New York; however, there were no days of "unsatisfactory" air quality during the sampling period. Haze was evident during the collection of Pomona sample No. 5, but was moderate relative to most smog episodes in the Basin area. Pomona sample No. 6 is a composite of four very similar days during which very little haze was evident. Pomona sample No. 7 is a composite of aerosol collected by EPA which was pooled for the purpose of obtaining an organic extraction large enough for replicate analyses performed to assess the reproducibility of the analytical procedure. The Denver sample was of interest because of its rather unusual and uncertain origin in the outlying regions of Denver. The Rubidoux and West Covina samples represent simultaneous collections at two points (about 50 miles apart) along a line extending due west from Los Angeles. Primary auto exhaust particulate was collected from a dilution tunnel in which exhausts were diluted during chassis dynamometer operations. Sample No. 12 is a pooled sample of secondary and primary auto exhaust particulate obtained by filtering the contents of a 17 m^3 smog chamber after irradiating diluted auto exhausts (8 ppm C) for 6 hours. The samples were collected as part of API Project EF-8 at Battelle-Columbus. Details regarding generation of the automotive samples are available.(36)

The data presented here serve to illustrate the types of numerical data that can be developed for the organic composition of particulate. In Appendix B the procedures employed are described in sufficient detail that other chemists may find them applicable in their work. Although the number of samples undertaken for organic analysis during this program was not sufficient to permit statistical analyses, it is hoped that further utilization of the described procedures will yield data that can be more fully analyzed and interpreted. Indeed, it is expected that apart from interpretations presented in this report, the data will be useful to others investigating the chemistry of atmospheric aerosols.

TABLE 6. SUMMARY OF AIR QUALITY CONDITIONS DURING ORGANIC PARTICULATE COLLECTIONS

Sample No.	Site	Date	Weather Conditions			Aerosol Mass Loading, $\mu g/m^3$				Light Scattering, 10^{-4} m^{-1}	Pollutants			
						Quartz,	Millipore				24-Hr Avg		1-Hr Max	
			General[b]	Temp, C	RH, %	<2 μm	A-Total	B<2 μm	B/A		THC, ppm	SO_2, ppm	NO_2, ppm	O_3, ppm
1	Columbus	July 21, 1972	S	30	64	69.5	85.9	72.9	0.85	5.8	3.4	0.020	0.148	0.117
2	Columbus	July 26, 1972	S	22	—	34.6	60	36.5	0.61	1.6	2.9	0.020	—	0.087
3	New York	August 23, 1972	PC	25	74	53.4	179.0	71.9	0.40	3.5	5.0	0.020	0.144	0.094
4	New York	August 11, 1972	PC	21	55	34.3	81.1	23.8	0.29	1.4	3.5	0.022	0.033	0.117
5	Pomona[a]	November 18, 1972	PC	—	—	52.4	114.2	55.0	0.48	2.4[a]	3.2	0.015	0.10	0.04
6	Pomona[a]	November 10-13, 1972	S	—	—	17.5	42.7	26.2	0.30	1.4[a]	2.0	0.011	0.082	0.03
7	Pomona	October 4-18, 1972	(Composite sample furnished by EPA for replication purposes only.)											
8	Denver[b]	November 16-17, 1973	—	4	40	—	—	—	—	2.8	3.5	0.011	0.04	0.04
9	Rubidoux[b]	September 21, 1973	—	—	—	—	—	—	—	—	—	—	—	—
10	West Covina[c]	September 21, 1973	PC	18	76	—	123	—	—	4.5	4.5	0.013	0.13	0.17

(a) Light-scattering data for Pomona were inferred from visibility observations (Bracket Field Airport, LaVerne, California). Pollutant data were supplied by the County of Los Angeles Air Pollution Control District.

(b) Data supplied by EPA.

(c) Data supplied by Battelle-Columbus (Project CAPA-9).

Solvent-Extractable Particulate Matter. As shown in Table 7, samples 1-6 were subjected to Soxhlet extraction, first with methylene chloride (20 hours) and then with dioxane (44 hours). Sequential Soxhlet extraction using methylene chloride followed by dioxane was used to obtain a wider range of organics than are extractable using the more commonly applied single-solvent procedures. Dioxane is expected to dissolve the polar, relatively oxidized organics normally left behind by benzene or chlorinated solvents. The data shown in Table 7 indicate that the dioxane extractables do indeed represent a significant fraction of the total extractable organic matter.

As noted in Appendix B, an experiment was conducted to verify that dioxane does not extract significant quantities of inorganic salt from the filters. Several salts were stirred individually with dioxane, and the filtered solvent was lyophilized in tared flasks. The data indicate that such inorganic salts do not correspond to >4 percent of the dioxane extractable matter, and that most likely this value is <1 percent of the dioxane-extractable matter.

Throughout the course of the determinations, solvent and filter blanks were carried through the reflux and concentration procedure. In calculating the values shown in Table 7, blanks due to solvent and filter background have been subtracted. Typically, three 182-cm^2 quartz-tissue filters were extracted using 100 ml of solvent. Blanks were determined as 0.09 mg (±0.01 mg) per filter for methylene chloride extractions, and 0.8 mg (±0.3 mg) per filter for dioxane extractions. The magnitude and variability of the dioxane blank is such that data concerning the dioxane-extractable matter should be interpreted cautiously. Nevertheless, it is felt that these values are not entirely spurious, and that they demonstrate that a significant fraction of atmospheric particulate matter consists of relatively oxidized organic matter. This assertion is supported by data for weight percent C, H, and N determined for methylene chloride and dioxane-extractable matter.

In view of the difficulties associated with the blanks where dioxane was used as an extraction solvent, atmospheric samples 7-10 and automotive samples 11 and 12 were extracted only with methylene chloride. Additionally, in an effort to further reduce blank values, all field sampling after 1972 was conducted using preextracted quartz-tissue filters. In these cases, preextraction was conducted first with methanol, then with benzene. Preextraction with methylene chloride led to a high chloride background. The preextracted material was shown to have a methylene chloride blank of 0.002 mg per filter and a methanol blank of 0.2 mg per filter. Thus, the reduction of background organic matter by preextracting was significant. Samples of primary and irradiated automotive particulate were collected on Gelman Type A glass-fiber filters (not preextracted). The blank for the Gelman filters was determined as 0.020 mg methylene chloride extractable matter per 81-cm^2 filter.

In considering the data in Table 7, note first the good agreement between replicate samples 7A and 7B. It is emphasized that this sample was divided after the extraction step. The methylene chloride extract was halved gravimetrically, after which concentration of the extract, determination of aliquot-weights (details, Appendix B), and all subsequent analytical procedures were conducted independently on the replicates. It was felt that cutting the filters in half initially to permit duplicate extractions would introduce an error of itself. Moreover, such division of filter samples is not a usual step in the scheme.

For 2 of 3 sampling sites (compare Samples 3-6, Table 7), it appears that a substantially larger fraction of dioxane extractable matter was present on days of more intense haze. Because highly oxidized materials are extracted by dioxane, a greater fraction of such compounds may be present on hazy days. The methylene chloride extractable fraction was fairly similar for all of

TABLE 7. SOLVENT EXTRACTION OF PARTICULATE MATTER

Sample No.	Site	Date	Mass Loading, $\mu g/m^3$	Total Particulate Taken for Extraction, mg	No. of Filter Disks Extracted	Methylene Chloride Extraction: Weight Extractable Matter, Corrected for Blank, mg[a]	Methylene Chloride Extraction: Weight Percent Extractable	Dioxane Extraction: Weight of Extractable Matter, Corrected for Blank, mg	Dioxane Extraction: Weight Percent Extractable
1	Columbus	July 21, 1972	69.5	128.6	2.5	10.58	8	10.91	8
2		July 26, 1972	34.6	63.1	2.5	10.15	16	17.45	28
3	New York	Aug. 11, 1972	34.3	55.6	2.25	9.17	16	5.38	10
4		Aug. 23, 1972	53.4	89.0	2.25	11.50	13	20.07	23
5	Pomona	Nov. 10-13, 1972	17.5	82.0	1.5	17.65	22	13.24	16
6		Nov. 18, 1972	52.4	60.0	1.5	9.52	16	24.03	40
7	Pomona	Oct. 4-18, 1972	—	465.1	4				
7A	Replicate A		—	—	—	29.60	13	11.5	5
7B	Replicate B		—	—	—	29.50	13	12.0	5
8	Denver	Nov. 16-17, 1973	—	26.56	2	12.03	45	—	—
9	Rubidoux	Sept. 21, 1973	—	167.5	1	15.40	9	—	—
10	West Covina	Sept. 21, 1973	—	47.64	1	7.29	15	—	—
11	Primary Automotive Exhaust		—	27.3	5	3.78	14	—	—
12	Irradiated Automotive Exhaust		—	19.7	26	2.26	11	—	—

(a) Sample Nos. 1-7: Circular quartz-tissue filters ($1.82 \times 10^{-2} m^2$), not preextracted, corrected for methylene chloride blank of 3.2×10^{-3} mg/m^2.
Sample Nos. 8-10: Rectangular quartz-tissue filters ($5.16 \times 10^{-2} m^2$), preextracted, corrected for methylene chloride blank of $.067 \times 10^{-3}$ mg/m^2.
Sample Nos. 11-12: Circular Gelman-A glass fiber filters ($8.1 \times 10^{-3} m^2$), not preextracted, corrected for methylene chloride blank of 1.6×10^{-3} mg/m^2.

the samples, with the outstanding exception of Denver, and showed no consistent trends with gross aerometric conditions.

Weight Percent C, H, and N. Values of weight percent C, H, N, and [O] for the methylene chloride and dioxane extractables are shown in Table 8. Determinations of weight percent C, H, and N were conducted using a Perkin-Elmer Model 240 elemental analyzer with gas purification accessory. The instrument performs automated Pregel-Dumas determinations. Values shown for weight percent oxygen [O] were calculated by difference from the CHN data. Approximately 1 mg of extractable matter was used for each determination. As expected, weight percent [O] is significantly higher for the dioxane extractables than for the methylene extractables. Within these fractions, however, there is little variation on extreme days, and not much variation from site to site, again the Denver sample may be an exception. Primary exhaust aerosol showed considerably less [O] and N in the methylene chloride fraction compared to the atmospheric samples. There was insufficient material in the sample of irradiated automotive aerosol to permit CHN analysis. As indicated in succeeding sections, however, infrared spectroscopic analysis and functional group analysis indicate that the irradiated automotive aerosol does indeed have a higher proportion of oxygen and nitrogen containing organics.

Infrared Spectra. Relative compositions of the methylene chloride and dioxane extractables were considered previously. Continuing along these lines, infrared spectra are shown in Figure 19 for methylene chloride and dioxane extractables of particulate collected in Pomona on November 18. Note that the absorbance due to C-H stretching, near 2900 cm^{-1}, is relatively strong for the methylene chloride extractables and relatively weak for the dioxane extractables. Absorbances near 1100 cm^{-1} may be assigned to C-O stretching for a variety of compounds including ethers, lactones, and esters. This region shows significantly stronger absorption with the dioxane extractables. Both samples show strong absorptions for carbonyl compounds in the region above 1700 cm^{-1}. These spectra indicate that the dioxane extractables consist of the relatively oxidized aerosol constituents compared with the less oxidized methylene chloride extractables.

Examination of the spectra compiled in Appendix C, indicates that the spectra of methylene chloride extractable matter generally exhibit significantly more detail and fine structure. In view of this and the difficulties associated with use of dioxane as an extraction solvent, this discussion will be confined to consideration of spectra for the methylene chloride extractable matter.

In order to present spectroscopic results in numerical form, key bands were selected and relative spectroscopic intensities were determined. Relative intensities were calculated by taking the ratio of optical density for the specified band to that observed for the CH stretching vibration at 2920 cm^{-1}; i.e.,

$$\text{Relative Intensity} = \frac{\text{(O.D. at specified absorption)}}{\text{(O.D. at 2920 cm}^{-1}\text{)}}$$

Such relative intensities have been calculated for the carbonyl band, the "percarbonyl" band (discussed below) and bands tentatively assigned as corresponding to organic nitrate.

TABLE 8. WEIGHT PERCENT CARBON, HYDROGEN, NITROGEN, AND OXYGEN IN SOLVENT EXTRACTABLE MATTER

Sample No.	Site	Date	Methylene Chloride Extractables, weight percent				Dioxane Extractables, weight percent			
			C	H	N	[O]	C	H	N	[O]
1	Columbus	July 21, 1972	69.3	9.5	0.8	20.4	44.4	6.1	0.8	48.7
2		July 26, 1972	73.0	10.2	0.9	15.9	47.2	6.0	0.8	46.0
3	New York	Aug. 11, 1972	68.8	9.0	1.3	20.9	46.6	5.5	2.2	45.7
4	"	Aug. 23, 1972	69.1	9.1	1.1	20.7	54.4	4.8	1.6	39.2
5	Pomona	Nov. 10-13, 1972	69.3	8.9	1.3	20.5	50.6	5.9	1.6	41.9
6		Nov. 18, 1972	61.6	7.7	1.9	28.8	39.1	5.0	1.9	54.0
7	Pomona	Oct. 4-18, 1972								
7A	Replicate A		63.2	7.7	1.4	27.7	–	–	–	–
7B	Replicate B		62.9	8.2	1.4	27.5	–	–	–	–
8	Denver	Nov. 16-17, 1973	76.5	10.5	0.8	12.2	–	–	–	–
9	Rubidoux	Sept. 21, 1973	60.6	7.9	2.2	29.3	–	–	–	–
10	West Covina	Sept. 21, 1973	65.7	8.6	1.5	24.2	–	–	–	–
11	Primary Automotive Exhaust		82.4	11.4	0.5	5.7	–	–	–	–

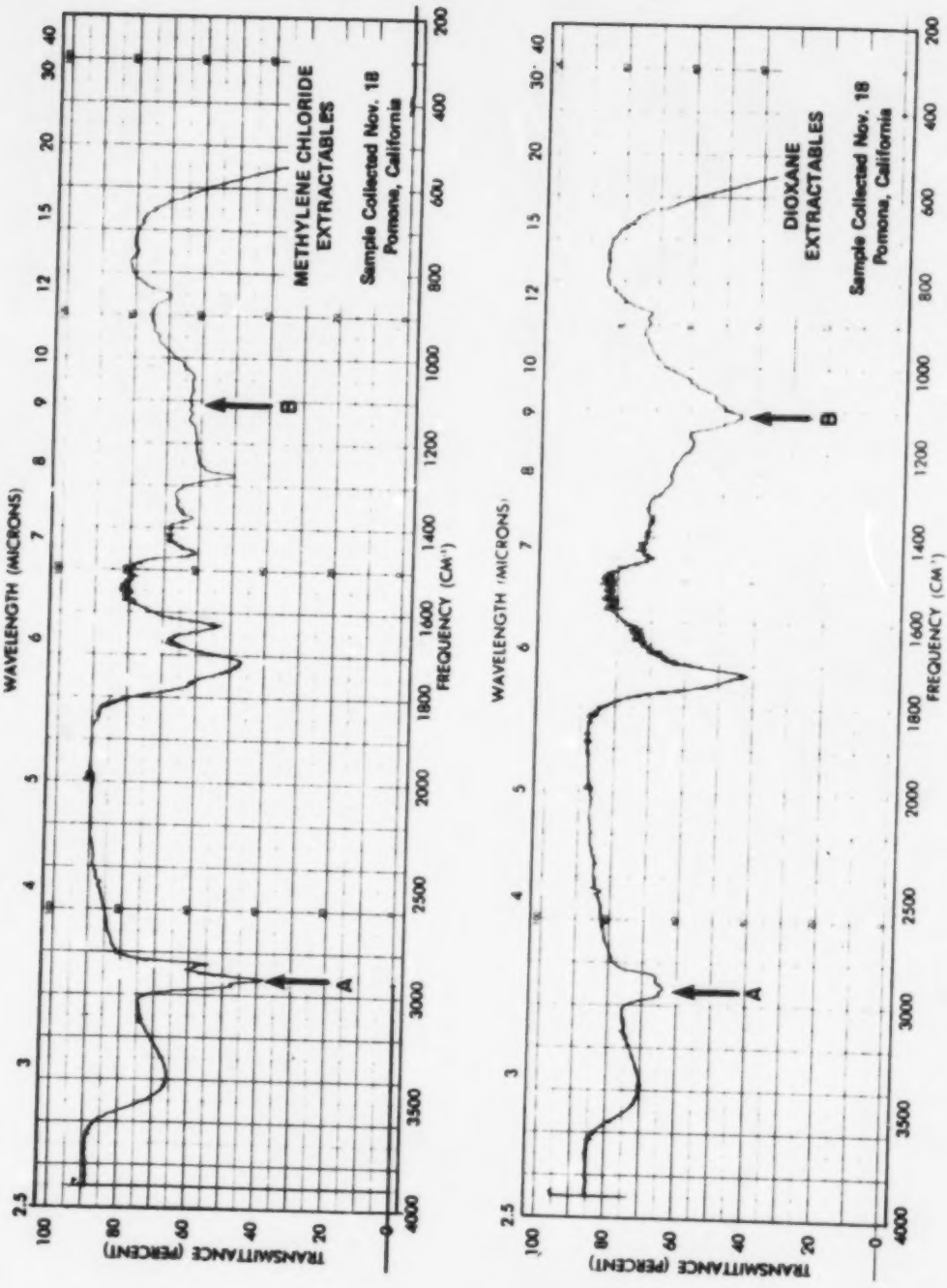
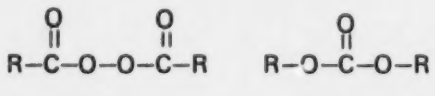


FIGURE 19. INFRARED SPECTRA OF SOLVENT-EXTRACTABLE MATTER

Group-frequency assignments in the range of 1700-1740 cm^{-1} have been made for a variety of carbonyl-containing compound types. The spectra of the methylene chloride extractables reveal some apparent trends in position and intensity of the carbonyl peak. A subtle variation in the position and width of the peak appears to be related to the weight percent oxygen in the sample. Samples with high weight percent oxygen display a subtle shift or broadening of the carbonyl peak to slightly higher frequency. This may be seen in the spectra shown in Figure 20 (carbonyl peaks are noted with Marker B). This effect is consistent with a greater population of diketones, keto acids, and keto aldehydes in the samples having higher oxygen incorporation. As shown in Table 8, the November 18 sample has 28.8 weight percent oxygen, while the November 10-13 sample has 20.5 weight percent oxygen.

Table 9 includes a listing of relative intensity of the carbonyl band as well as the value for weight percent oxygen in the methylene chloride extractable matter. Examination of these data reveals a good correlation (R = 0.90) between oxygen incorporation and carbonyl band intensity.

A second interesting effect concerns a shoulder near 1770 cm^{-1} (Marker A in Figure 20). The shoulder may be assigned to peroxides, organic carbonates, anhydrides or lactones. It is most probably for alkyl/aryl peroxides and carbonates. A relatively limited variety of anhydrides or lactones would exhibit this absorption, principally five- or six-membered strained cyclic species. Thus, this absorption is probably due to peroxides or carbonates. Since this absorption cannot be unambiguously assigned, we will refer to it as the "percarbonyl" peak. This is meant to refer to the relatively oxidized character of the candidate compound types, especially the peroxides and carbonates.



Peroxide Carbonate

Regarding both the carbonyl and percarbonyl bands, good reproducibility is apparent for samples 7A and 7B. There is a substantial difference in band intensities for the primary and irradiated automotive exhaust aerosols. As expected, the irradiated sample shows greater content of carbonyl and percarbonyl compound types, and the band intensities for the irradiated sample are of comparable magnitude to those observed for the more highly oxidized atmospheric samples.

Interesting trends in peak intensities were also observed for absorptions at 1630 cm^{-1} and 1275 cm^{-1} (Figure 20, Markers C and D). According to Colthup, et al.[37], organic nitrates ($R-O-NO_2$) display strong bands near 1660 to 1625 cm^{-1} (NO_2 symmetric stretching). Broad absorptions observed at 850 cm^{-1} are also consistent with organic nitrates. The observed bands, however, might also be attributed to aromatic amines. Nevertheless, the group frequencies for aromatic amines are quite broad. For example, the band due to CN stretching could appear in the region of 1250 to 1380 cm^{-1}. Most primary amines have an NH_2 deformation band at 1650 to 1590 cm^{-1}, while secondary aromatic amines have an NH bending absorption near 1510 cm^{-1}. Most aliphatic secondary amines have no significant NH bending band above 1470 cm^{-1}. Thus, although various aromatic amines could give rise to the observed bands at 1275 cm^{-1} and 1630 cm^{-1}, these absorptions are more generally assignable to organic nitrates. The observed bands are therefore tentatively assigned to organic nitrates. Values for relative band intensity are shown in Table 10 along with data for weight percent nitrogen. A trend in the

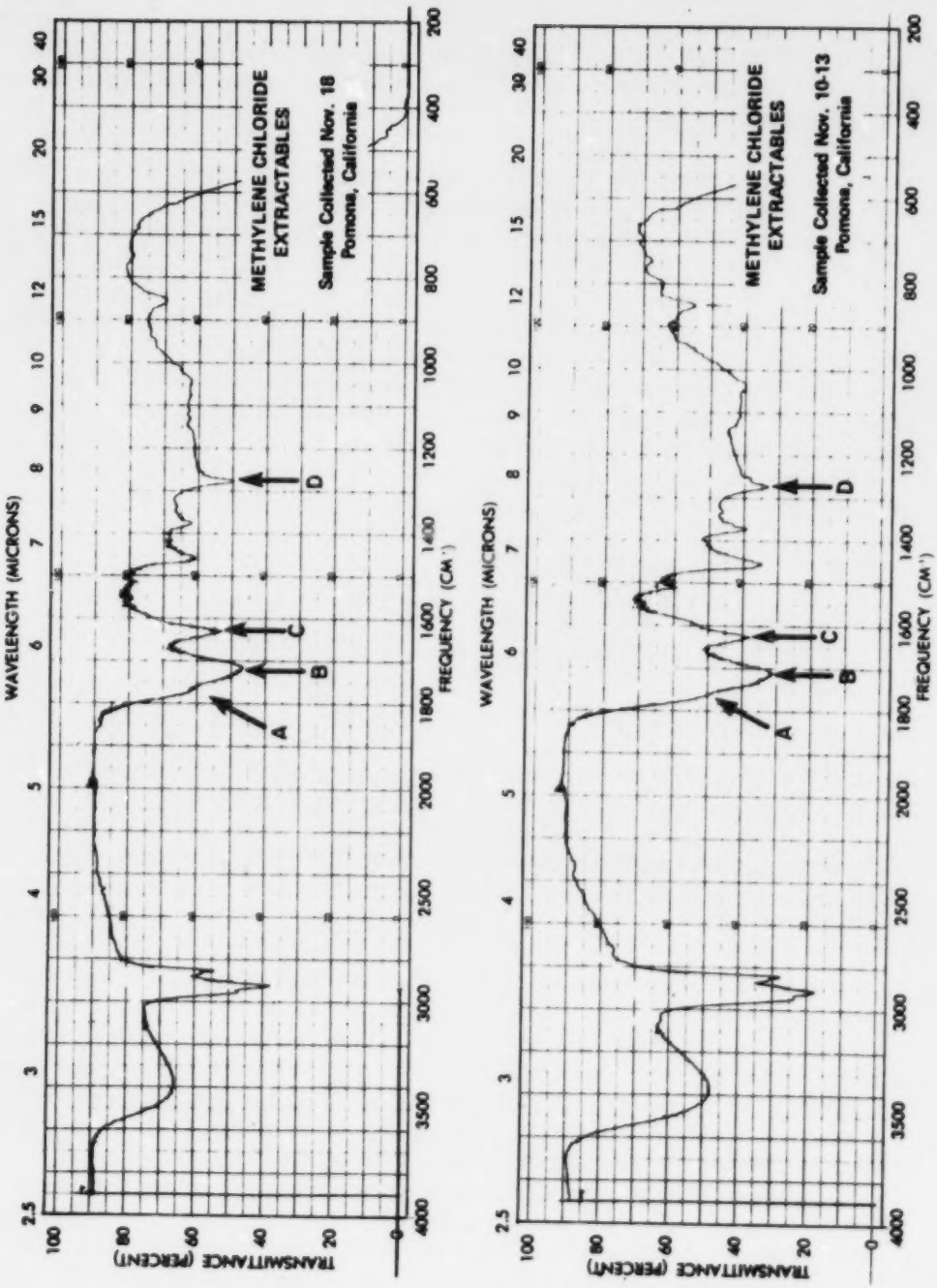


FIGURE 20. INFRARED SPECTRA OF SOLVENT-EXTRACTABLE MATTER

TABLE 9. INFRARED SPECTROSCOPIC DATA ON CARBONYL BANDS

Data presented are for methylene chloride extracts.

Sample No.	Site	Date	Weight percent [0]	Carbonyl Band Relative Intensity	Percarbonyl Band[a] Relative Intensity
1	Columbus	July 21, 1972	20.4	0.58	0.16
2		July 26, 1972	15.9	0.38	0.07
3	New York	Aug. 23, 1972	20.7	0.47	0.14
4		Aug. 11, 1972	20.9	0.50	0.13
5	Pomona	Nov. 18, 1972	28.8	0.64	0.06
6		Nov. 10-13, 1972	20.5	0.51	0.03
7	Pomona	Oct. 4-18, 1972			
7A	Replicate A		27.7	0.56	0.05
7B	Replicate B		27.5	0.56	0.05
8	Denver	Nov. 16-17, 1973	12.2	0.13	<0.02
9	Rubidoux	Sept. 21, 1973	29.3	0.96	0.16
10	West Covina	Sept. 21, 1973	24.2	0.64	0.12
11	Primary Automotive Exhaust		5.7	0.06	<0.02
12	Irradiated Automotive Exhaust		—	0.77	0.16

(a) Discussed in text.

TABLE 10. INFRARED SPECTROSCOPIC DATA ON NITRATE BANDS

Data presented are for methylene chloride extracts.

Sample No.	Site	Date	Weight Percent N	Nitrate Band I[(a)] (1630 cm^{-1}) Relative Intensity	Nitrate Band II[(a)] (1275 cm^{-1}) Relative Intensity	Average Intensity, Bands I and II
1	Columbus	July 21, 1972	0.8	0.17	0.19	0.18
2		July 26, 1972	0.9	0.15	0.15	0.15
3	New York	Aug. 11, 1972	1.3	0.25	0.28	0.27
4		Aug. 23, 1972	1.1	0.19	0.22	0.21
5	Pomona	Nov. 18, 1972	1.9	0.43	0.37	0.40
6		Nov. 10-13, 1972	1.3	0.29	0.24	0.27
7	Pomona	Oct. 4-18, 1972				
7A	Replicate A		1.4	0.22	0.25	0.24
7B	Replicate B		1.4	0.21	0.26	0.24
8	Denver	Nov. 16-17, 1973	0.8	0.04	0.05	0.05
9	Rubidoux	Sept. 21, 1973	2.2	0.47	0.47	0.47
10	West Covina	Sept. 21, 1973	1.5	0.33	0.32	0.33
11	Primary Automotive Exhaust		0.5	<0.02	<0.02	<0.02
12	Irradiated Automotive Exhaust		—	0.34	0.49	0.42

(a) Tentative assignments.

data is clear, and correlation coefficients for nitrate bands I and II with nitrogen content are 0.87 and 0.89, respectively.

Note in particular the enhanced intensity of nitrate peaks in the irradiated automotive aerosol compared to the primary automotive aerosol. Nitrate band intensities for the irradiated aerosol are comparable to those determined for atmospheric samples having the greatest degree of nitrogen incorporation. Again, good reproducibility was obtained for band intensities in replicate samples 7A and 7B.

Aromatic/Aliphatic Ratio. Nuclear magnetic resonance spectra were obtained using a 60 MHz Varian Associates spectrometer equipped to perform Fourier-transform spectroscopy. Fourier-transform spectroscopy permits the rapid generation of spectra, consequently allowing the generation of a large number of spectra in a relatively short period of time. Computer time-averaging a series of such spectra leads to enhanced signal-to-noise ratios and permits acquisition of useful spectra not obtainable by conventional techniques. The current application required the use of FT-NMR to obtain adequate spectra. Spectra were obtained using ~1-mg samples of extractable matter. Methylene chloride extractables were analyzed in deutero-chloroform solution and dioxane extractables were analyzed in deutero-dioxane solution. From the spectra obtained, aromatic/aliphatic ratios were computed. The values were calculated on the basis of the integrated resonances for methyl, methylene and aromatic protons; i.e.,

$$\text{Aromatic/Aliphatic Ratio} = \frac{\text{(integrated resonances; aromatic protons)}}{\text{(integrated resonances; methyl + methylene protons)}}$$

The results are shown in Table 11. Results for the dioxane extractable matter have been omitted. It was observed that upon standing, a fine precipitate formed in the concentrated deutero-dioxane solutions leading to irregular results. Data for automotive aerosol samples have been omitted because of sample limitations. Good reproducibility was obtained for replicate samples 7A and 7B. Reasons for the order of magnitude difference in aromatic/aliphatic ratios are not apparent. It is clear, however, that aliphatic protons are dominant in the methylene chloride extractions.

Acid/Base Neutral Distribution. The methylene chloride extractable matter was fractionated according to the scheme shown below. The material was fractionated first into water-soluble and water-insoluble components by partitioning the methylene chloride solution against distilled water. Water was removed by lyophilization, and the weight of the water-soluble fraction was determined.

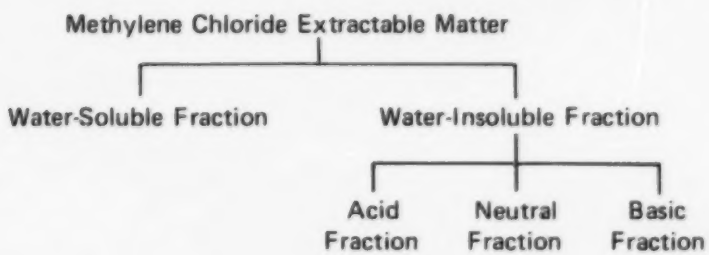


TABLE 11. FOURIER TRANSFORM-NMR ANALYSIS OF METHYLENE CHLORIDE EXTRACTABLES

Sample No.	Site	Date	Aromatic/ Aliphatic Ratio
1	Columbus	July 21, 1972	0.10
2		July 26, 1972	0.11
3	New York	Aug. 23, 1972	0.12
4		Aug. 11, 1972	<0.01
5	Pomona	Nov. 18, 1972	0.15
6		Nov. 10-13, 1972	<0.01
7	Pomona	Oct. 14-18, 1972	
7A	Replicate A		0.15
7B	Replicate B		0.14
8	Denver	Nov. 16-17, 1973	0.04
9	Rubidoux	Sept. 21, 1973	<0.01
10	West Covina	Sept. 21, 1973	0.01

The water-insoluble material remaining in methylene chloride solution was next fractionated into acid, basic, and neutral components. The solution was extracted using first 2N aqueous sodium hydroxide and then 2N hydrochloric acid. Material remaining in methylene chloride solution after extractions with both aqueous sodium hydroxide and hydrochloric acid is defined as the water-insoluble neutral fraction.

The sodium hydroxide extract (containing organic-acid salts) was brought to pH ~0.8, and the free acid was extracted into methylene chloride, using a continuous liquid/liquid extractor. Similarly the hydrochloric acid extract (containing organic-base salts) was brought to pH ~13, and the free base was extracted into methylene chloride, again using a continuous liquid/liquid extractor. The methylene chloride solutions containing the acid, basic, and neutral fractions were dried by refluxing the solutions over a 3 A molecular sieve. The dried solutions were concentrated and small aliquots of the concentrate were evaporated to dryness to determine the residue weights for the various fractions. Using these data, values for the weight percent distribution were calculated.

Details of the fractionation procedure are presented in Appendix B. The procedure represents a departure from methods most often employed in such fractionation. To minimize sample losses to the point where meaningful distribution data could be reported for the small samples available, special techniques including vapor-phase drying and continuous liquid/liquid extraction were incorporated in the procedure. As refinement of the scheme progressed, however, it became clear that unacceptable material losses had occurred during fractionation of some of the earlier samples, notably samples 1, 2, 5, and 6. Thus, data for these samples has not been reported.

Data shown in Table 12 were calculated on the basis of recovered material. Thus the distribution data totals 100 percent for each sample. Overall recoveries varied from 60 to 100 percent. Typically, about 5 mg of sample was subjected to fractionation. Wet chemical analysis of small samples inevitably results in some random handling losses. The recovery observed here is consistent with the quantity of sample used and the nature of the analytical procedure applied. The reproducibility between replicate samples 7A and 7B is good.

Partitioning of the methylene chloride fraction of the two samples from New York indicates similar distributions in spite of different atmospheric conditions. Samples of irradiated auto exhaust and those from Rubidoux and West Covina (but not Pomona) have relatively high water-soluble fractions. Again, the Denver sample seems somewhat inconsistent with the other atmospheric samples, particularly within the water-insoluble fraction where higher neutral and lower acidic fractions are noted. Among the water-insoluble fractions (of methylene chloride extract) the neutral fraction accounts for 60-70 percent of the mass and the acids about 30-40 percent.

Functional Group Analyses. Quantitative functional group analyses for alcohol and for carbonyl were performed using the water-insoluble neutral fraction of the methylene chloride extractable matter. Isolation of this fraction was considered above. Details of the fractionation and functional group procedures are presented in Appendix B.

The procedure for determination of alcohols is subject to interference from carboxylic acids, phenols, and amines. Thus, the analysis is conducted only on the neutral fraction. Although the carbonyl determination is not subject to the interferences that limit the alcohol analysis, the

TABLE 12. FRACTIONATION OF METHYLENE CHLORIDE EXTRACTABLE MATTER

Sample No.	Site	Date	Mass Distribution, weight percent methylene chloride extract			
			Water-Insoluble Fraction			Water-Soluble Fraction
			Acid	Neutral	Base	Total
3	New York	Aug. 11	26	52	4	18
4		Aug. 23	30	50	5	15
7	Pomona	Oct. 4-18				
7A	Replicate A		24	48	<1	28
7B	Replicate B		22	51	<1	26
8	Denver	Nov. 16-17, 1973	18	66	3	14
9	Rubidoux	Sept. 21, 1973	24	33	1	42
10	West Covina	Sept. 21, 1973	20	36	2	42
12	Irradiated Automotive Exhaust		12	33	1	54

carbonyl determination is applied to the neutral fraction in order that a comparison can be made between alcohol and carbonyl content in the same sample fraction.

Results of functional group analyses were obtained on molar bases. This is well suited to the current application, in which quantitative measure of specified organic functionalities is desired. However, to permit a direct comparison between weight percent carbonyl oxygen, weight percent alcoholic oxygen, and weight percent oxygen obtained from the CHN determination performed on the unfractionated extract, the molar values for carbonyl and alcohol were converted to a common basis of weight percent oxygen. This was calculated as shown in the following example:

$$\text{Weight percent carbonyl oxygen} = \frac{(\text{mg [O] present as carbonyl})}{(\text{mg sample})} \times 100.$$

The data are shown in Table 13. In considering the results of the CHN determinations, it was noted that incorporation of oxygen into organic particulate may be a useful indicator of the tendency for a given atmosphere to oxidize gas-phase or suspended organics. In terms of fundamental atmospheric reactions, incorporation of oxygen to yield carbonyl compounds might be related to ozonolysis, while incorporation to yield alcohols might be related to free-radical processes.

With the exceptions of the Columbus No. 1 and the Denver No. 8 samples, which were higher and lower, respectively, the alcoholic oxygen fraction of the neutral samples were fairly constant. The irradiated automotive sample showed a much higher content in this category. The fractional carbonyl content was similar for Columbus and New York samples with somewhat higher percentages occurring for Denver, Rubidoux, West Covina, and particularly irradiated automobile exhaust.

Further examination of the data does not reveal meaningful correlation between functional group concentrations and weight percent oxygen in the methylene chloride extract or functional group concentrations and average ozone concentration on the respective sampling days. Indeed, the values for carbonyl and total oxygen concentrations do not correlate with each other.

Summary. Organic fractionation and functionality data are summarized in Table 14 for aerosol collected on days of substantial light scattering at four sites and from smog chamber irradiations of auto exhaust. The data are presented on a weight percent basis and therefore reflect compositional differences and/or similarities among the samples.

Comparing first the amounts of methylene chloride extractable material, it is apparent that extractable fractions are quite similar in all cases except Denver. The dioxane extractable fraction, not included here, probably accounts for another ~20 percent of the particulate mass. (Note that all data in Table 14 pertain to the methylene chloride extractable fractions). Reviewing the CHN[O] data, it appears that the Denver sample is composed of more saturated matter (higher C-H values) than the other samples and that the West Coast and irradiated exhaust samples consist of higher percentages of [O] and N than the Denver and New York samples. Partitioning the methylene chloride fraction resulted in identical water-soluble and insoluble fractions of the West Covina and Rubidoux samples, which are somewhat similar to the analyses on the irradiated exhaust samples and dissimilar to those on the Denver and New York samples. The basic fraction of the water-insoluble material was small for all samples and, although some

TABLE 13. FUNCTIONAL GROUP ANALYSIS OF METHYLENE CHLORIDE NEUTRAL FRACTION

Sample No.	Site	Date	Weight Percent [O] $MeCl_2$ Extract	Alcoholic Oxygen, weight percent in neutral fraction	Carbonyl Oxygen, weight percent in neutral fraction	Total Oxygen, Alcoholic Plus Carbonyl, weight percent
1	Columbus	July 21, 1972	20.4	19.3	4.9	24.2
2		July 26, 1972	15.9	5.0	5.5	10.5
3	New York	Aug. 23, 1972	20.7	3.0	4.2	7.2
4		Aug. 11, 1972	20.9	4.6	5.2	9.8
5	Pomona,	Nov. 18, 1972	28.8	4.0	7.3	11.3
6		Nov. 10-13, 1972	20.5	4.4	3.2	7.6
7	Pomona	Oct. 4-18, 1972				
7A	Replicate A		27.7	1.7	4.5	6.2
7B	Replicate B		27.5	1.7	4.4	6.1
8	Denver	Nov. 16-17, 1973	12.2	0.5	13.1	13.6
9	Rubidoux	Sept. 21, 1973	29.3	5.0	11.1	16.1
10	West Covina	Sept. 21, 1973	24.2	3.5	12.1	15.6
12	Irradiated Automotive Exhaust		–	33	21	54

BATTELLE - COLUMBUS

TABLE 14. SUMMARY OF METHYLENE CHLORIDE FRACTIONATIONS

Site: / Date:	New York Aug. 23, 1972	Denver Nov. 17, 1973	Rubidoux Sept. 21, 1973	W. Covina Sept. 21, 1973	Irradiated Auto Exhaust
Methylene Chloride Extractables, weight percent	13	45	9	15	11
Carbon, weight percent	69.1	76.5	60.6	65.7	61.6
Hydrogen, weight percent	9.1	10.5	7.9	8.6	7.7
Nitrogen, weight percent	1.1	0.8	2.2	1.5	1.9
[Oxygen], weight percent	20.7	12.2	29.3	24.2	28.8
Fractionation					
Water-soluble fraction, weight percent	15	14	42	42	54
Water-insoluble fraction, weight percent	85	86	58	58	46
Acid fraction, weight percent	30	18	24	20	12
Basic fraction, weight percent	5	3	1	2	1
Neutral fraction, weight percent	50	65	33	36	33
Alcohol fraction, weight percent [O]	1.5	0.3	1.7	1.3	10.9
Carbonyl fraction, weight percent [O]	2.1	8.5	3.6	4.3	6.9

similarities can be seen in the remainder of these data, no significant pattern emerges, particularly in comparisons with the average auto-exhaust sample. In the last set of data on functional groups, the Denver and auto exhaust samples stand out as irregular – compared to the urban collections the Denver sample is low in alcohol and high in carbonyl constituents while the exhaust sample is high in both alcohol and carbonyl fractions.

Trends in these data are consistent with those from the spectroscopic analyses. In general, the West Coast samples (Rubidoux and West Covina) showed somewhat higher carbonyl concentrations. And, in accord with the total N determinations, the nitrate band intensities for Rubidoux, West Covina, and irradiated auto exhaust were greater than for Denver and New York.

In establishing the overall characteristics of the organic composition of light-scattering aerosols, we were looking for gross compositional differences. We found only slight and moderate differences among samples from different regions, with the exception of the unusual Denver sample. There were also only moderate organic compositional differences among days of different air quality in the same region. Among the trends is an expected one that a higher percentage of oxygenated material is present on days of higher light scattering.

If it is desirable to arrive at some tentative, undefendable conclusion with respect to relationships between auto exhaust emissions and the organic fraction of light-scattering aerosols, one might state that the West Coast samples showed some significant similarities to the composition of the auto exhaust samples, the New York samples showed marginal similarities, and the rural Denver samples showed no similarities. It must be stressed that only a very limited number of aerosol samples have been analyzed for organic constituency. It is our hope that the knowledge gained here, when combined with continuing work on the analysis of both smog chamber (model systems) and atmospheric aerosols, will eventually shed more light on the chemistry of organic aerosol formation.

REFERENCES

(1) Wilson, W. E., Schwartz, W. E., and Kinzer, G. W., "Haze Formation: Its Nature and Origin", First Year Report by Battelle's Columbus Laboratories to the Coordinating Research Council and the Environmental Protection Agency (1972).

(2) Miller, D. F., Schwartz, W. E., Jones, P. W., Joseph, D. W., Spicer, C. W., Riggle, C. J., and Levy, A., "Haze Formation: Its Nature and Origin", EPA-650-3-74-002, NERC, Research Triangle Park, N.C. (June, 1973).

(3) McNulty, R. P., Atmospheric Environment, 2, 625 (1968).

(4) Horvath, H., Atmospheric Environment, 5, 177 (1971).

(5) Beuttell, R. G., and Brewer, A. W., J. Sci. Instr., 26, 357 (1949).

(6) Ahlquist, N. C., and Charlson, R. J., Environ. Sci. and Technol., 3, 363 (1968).

(7) Samuels, H. J., Twiss, S., and Wong, E. W., "Visibility, Light Scattering and Mass Concentration of Particulate Matter", Report of the California Tri-City Aerosol Sampling Project to the State of California Air Resources Board (July, 1973).

(8) Charlson, R. J., Ahlquist, N. C., Selvidge, H., and McCready, P. B., J. Air Poll. Control Assoc., 19, 937 (1969).

(9) Buchan, W. E., and Charlson, R. J., Science, 159, 193 (1968).

(10) Covert, D. S., Charlson, R. J., and Ahlquist, N. C., J. Appl. Meteor., 11, 968 (1972).

(11) Lundgren, D. A., "Atmospheric Aerosol Composition and Concentration as a Function of Particle Size and Time", paper no. 69-128 presented at the 62nd Annual Air-Poll. Control Assoc. Meeting, New York City (1969).

(12) Whitby, K. T., Husar, R. B., and Liu, B.Y.H., J. Colloid Interface Sci., 39, 177 (1972).

(13) Hidy, G. M., "Theory of Formation and Properties of Photochemical Aerosols", paper presented at the Battelle School on Air Pollution, Seattle, Wash. (1973).

(14) Blosser, E. R., "A Study of the Nature of the Chemical Characterisitcs of Particulates Collected from Ambient Air", final report from Battelle's Columbus Laboratories to the National Air Pollution Control Administration (1970).

(15) John, W., Kaifer, R., Rahn, K., and Wesolowski, J. J., Atmospheric Environ., 7, 107 (1973).

(16) Lee, R. E., Goranson, S. S., Enrione, R. E., and Morgan, G. B., Environ. Sci. and Technol., 6, 1025 (1972).

(17) Winchester, J. D., and Nifong, G. D., Water, Air, and Soil Pollution, 1, 50 (1971).

(18) Gladney, E. S., Zoller, W. H., Jones, A. G., and Gordon, G. E., Environ. Sci. and Technol., 6, 551 (1974).

(19) Lee, R. E., and Patterson, R. K., Atmospheric Environ., 3, 249 (1969).

(20) Wagman, J., Lee, R. E., and Axt , C. J., Atmospheric Environ., 1, 479 (1967).

(21) Miller, M. S., Friedlander, S. K., and Hidy, G. M., J. Colloid and Interface Sci., 39, 165 (1972).

(22) Friedlander, S. K., Environ. Sci. and Technol., 7, 235 (1973).

(23) Heisler, S. L., Friedlander, S. K., and Husar, R. B., Atmospheric Environ., 7, 633 (1973).

(24) Novakov, T., Mueller, P. K., Alcocer, A. E., and Otvos, J. W., J. Colloid and Interface Sci., 39, 225 (1972).

(25) U. S. Department of Health, Education, and Welfare, "Air Quality Data, 1966", Durham, N.C. (1968).

(26) Hauser, T. R., and Pattison, J. N., Environ. Sci. and Technol., 6, 549 (1972).

(27) Ciaccio, L. L., Rubino, R. L., and Flores, J., Environ. Sci. and Technol., 10, 935 (1974).

(28) Mader, P. P., McPhee, R. D., Lofberg, R. T., and Larson, G. P., Industrial and Engineering Chem., 44, 1352 (1952).

(29) Gemma, J. L., and Miller, D. F., "A Model of Urban Visibility Based on Air Quality", presented at the Annual American Chemical Society Meeting, Atlantic City, N.J. (1974).

(30) Air Pollution, Edited by A. C. Stern, Second Edition, Vol. 1, Academic Press, New York (1968), "Meteorology and Air Pollution" (R.C. Wanta) 187-224.

(31) Sonquist, J. A., and Morgan, J. N., The Detection of Interaction Effects, SRC Monograph No. 35, University of Michigan Institute for Social Research, Ann Arbor, Mich. (1964).

(32) Box, G.E.P., and Jenkins, G. M., Time Series Analysis; Forecasting and Control, Holden-Day, San Francisco, Ca. (1970).

(33) Mueller, P. K., Mosley, R. W., and Pierce, L. B., J. Colloid and Interface Sci., 39, 235 (1972).

(34) Spicer, C. W., "Fate of Nitrogen Oxides in the Atmosphere", Final Report from Battelle's Columbus Laboratories to the Coordinating Research Council and the Environmental Protection Agency, Project CAPA-9 (1974).

(35) Altshuller, A. P., Environ. Sci. and Technol., 7, 709 (1973).

(36) Levy, A., Miller, D. F., Hopper, D. R., Spicer, C. W., and Trayser, D. A., "Motor Fuel Composition and Photochemical Smog", Interim Report from Battelle's Columbus Laboratories to the American Petroleum Institute, API Report No. CEA-4 (1973).

(37) Cothup, N. B., Daly, L. H., and Wiberly, S. E., Introduction to Infrared and Roman Spectroscopy, Academic Press, New York (1964).

APPENDIX A

STATISTICAL METHODS

AID Analysis

The procedure followed in AID is to divide the sample into a set of mutually exclusive and exhaustive subgroups, through a succession of binary splits of the data. At each step, the split is based on values of one predictor variable. The predictor variable, and its values associated with each subgroup of the split, are chosen in such a way as to minimize the error sum of squares of the dependent variable. The final result of applying the AID procedure is a classification tree, which yields a sequential decision rule based on predictor variables. The result of each "path" of the decision rule is a final subgroup. Since each final subgroup is characterized by the mean and standard deviation of dependent variable values included in it, and since the paths leading to final subgroups can be characterized by a sequence of sets of values of predictor variables, a quantitative statement can be formulated for each path and used for purposes of classification or prediction.

AID has advantages over traditional statistical methods such as regression analysis. With AID, no assumptions about linearity or homoscedasticity need be made. Interactions among predictor variables relative to a dependent variable do not invalidate an AID analysis. In fact, AID helps to identify such interactions in the data.

The following is a mathematical description of the AID algorithm.

1. The total input sample is considered the first (and indeed only) group at the start.

2. Select that unsplit sample group, group i, which has the largest total sum of squares

$$TSS_i = \sum_{\alpha=1}^{N_i} Y^2 - \frac{\left(\sum_{\alpha=1}^{N_i} Y_\alpha \right)^2}{N_i} \quad , \qquad (1)$$

such that for the i'th group

$$TSS_i \geq R\ (TSS_T) \text{ and } N_i \geq M \quad , \qquad (2)$$

where R is an arbitrary parameter (normally $.01 \leq R \leq .10$) and M is an arbitrary integer (normally $20 \leq M \leq 40$), and TSS_T denotes the total sum of squares for the total input sample.

The requirement (2) is made to prevent groups with little variation in them, or small numbers of observations, or both, from being split. That group with the largest total sum of squares (around its own mean) is selected, provided that this quantity is larger than a specified fraction of the original total sum of squares (around the grand mean), and that this group contains more than some minimum number of cases (so that any further splits will

be credible and have some sampling stability as well as reducing the error variance in the sample).

3. Find the division of the C_k classes of any single predictor X_k such that combining classes to form the partition p of this group i into two nonoverlapping subgroups on this basis provides the largest reduction in the unexplained sum of squares. Thus, choose a partition so as to maximize the expression

$$(n_1\bar{y}_1^2 + n_2\bar{y}_2^2) - N_i\bar{Y}_i^2 = BSS_{ikp} \quad , \qquad (3)$$

where $N_i = n_1 + n_2$

and $\bar{Y}_i = \dfrac{n_1\bar{y}_1 + n_2\bar{y}_2}{N_i}$

for group i over all possible binary splits on all predictors, with restrictions that (a) the classes of each predictor are ordered into descending sequence, using their means as a key and (b) observations belonging to classes which are not contiguous (after sorting) are not placed together in one of the new groups to be formed. Restriction (a) may be removed, by option, for any predictor X_k.

4. For a partition p on variable k over group i to take place after the completion of step 3, it is required that

$$BSS_{ikp} \geq Q\,(TSS_T) \quad , \qquad (4)$$

where Q is an arbitrary parameter in the range $.001 \leq Q < R$, and TSS_T is the total sum of squares for the input sample. Otherwise group i is not capable of being split; that is, no variable is "useful" in reducing the predictive error in this group. The next most promising group (TSS_j = maximum) is selected via step 2 and step 3 is then applied to it, etc.

5. If there are no more unsplit groups such that requirement (2) is met, or if, for those groups meeting it, requirement (4) is not met (i.e., there is no "useful" predictor), or if the number of currently unsplit groups exceeds a specified input parameter, the process terminates.

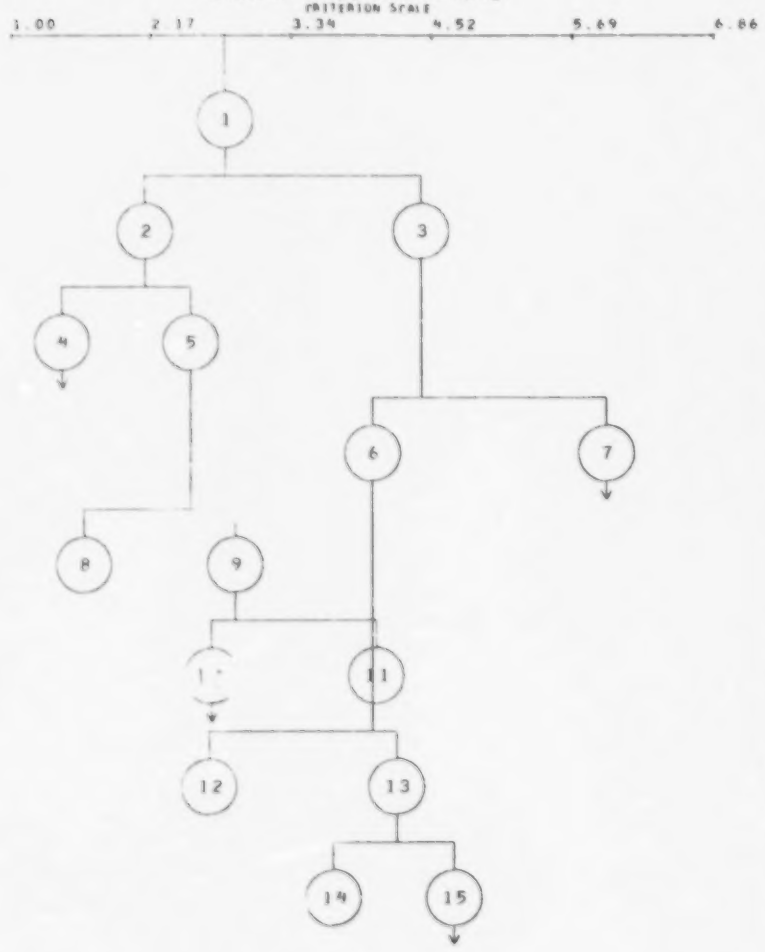
Figures A-1, A-2, and A-3 show the results obtained from AID analyses with light scattering as dependent variable. In order to understand the results, a description of how to analyze the output is included here. The criterion scale on the upper left of Figure A-1 is in units of the dependent variable, light scattering. The tabular information on the right begins with descriptive statistics about the sample. The criterion variable is given (LIGHT SCAT), there are 432 observations in the sample (18 days with 24 hourly averages of data each), and the average value of light scattering in this sample is 2.79 with a standard deviation of 1.58. The disk containing the number one represents this sample on the graph, and is located at the mean value (2.79) of the light-scattering observations in the sample.

The sample is split first by NO_2 into subgroups 2 and 3. On the graph, the subgroups are located at the mean value of the light-scattering observations in each subgroup. Information pertaining to this split is contained in the table immediately to the right of the graphical display. Hence, subgroup 2 contains 305 observations with an average value of light scattering of 2.11

AUTOMATIC INTERACTION DETECTOR

HAZEVECTOR ..... AID ANALYSIS

BINARY TREE STRUCTURE



SUMMARY TABLE

---TOTAL GROUP---

CRITERION - LIGHT SCAT

TOTAL GROUP N = 432
MEAN = 2.79
STD. DEV. = 1.58

PARENT 1 SPLITTING VARIABLE - NO2

MEAN= 2.11 S.D.= 1.10 N= 305
PREDICTOR VALUES -- 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

MEAN= 4.42 S.D.= 1.36 N= 127
PREDICTOR VALUES -- 16 17 18 19 20 21 22 23 24 25 26 27 28 29 31 32 33 34 35 37 38 39 40 41 42 43 44 47 .

PARENT 2 SPLITTING VARIABLE - TEMP

MEAN= 1.41 S.D.= 0.85 N= 106
PREDICTOR VALUES -- 0 1 2 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

MEAN= 2.49 S.D.= 1.03 N= 199
PREDICTOR VALUES -- 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 .

PARENT 3 SPLITTING VARIABLE - CO

MEAN= 4.01 S.D.= 1.11 N= 100
PREDICTOR VALUES -- 2 3 4 6 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 .

MEAN= 5.56 S.D.= 1.05 N= 27
PREDICTOR VALUES -- 38 39 40 41 42 43 44 45 46 48 50 52 54 55 57 58 59 61 63

PARENT 5 SPLITTING VARIABLE - HUMIDITY

MEAN= 1.59 S.D.= 0.51 N= 58
PREDICTOR VALUES -- 1 2 7 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
FINAL GROUP

MEAN= 2.84 S.D.= 0.97 N= 141
PREDICTOR VALUES -- 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 48 49 52 53 55 56 .

PARENT 9 SPLITTING VARIABLE - METH

MEAN= 2.66 S.D.= 0.80 N= 121
PREDICTOR VALUES -- 0 1 2 3 4 5 6 7 8 9 10 12 14 15 16 17 18 19 20 21 22 23 24

MEAN= 4.04 S.D.= 1.04 N= 20
PREDICTOR VALUES -- 25 26 27 28 29 30 31 33 34 37 44
FINAL GROUP

PARENT 6 SPLITTING VARIABLE - CO

MEAN= 2.64 S.D.= 0.79 N= 13
PREDICTOR VALUES -- 2 3 4 6 8 9 10 11
FINAL GROUP

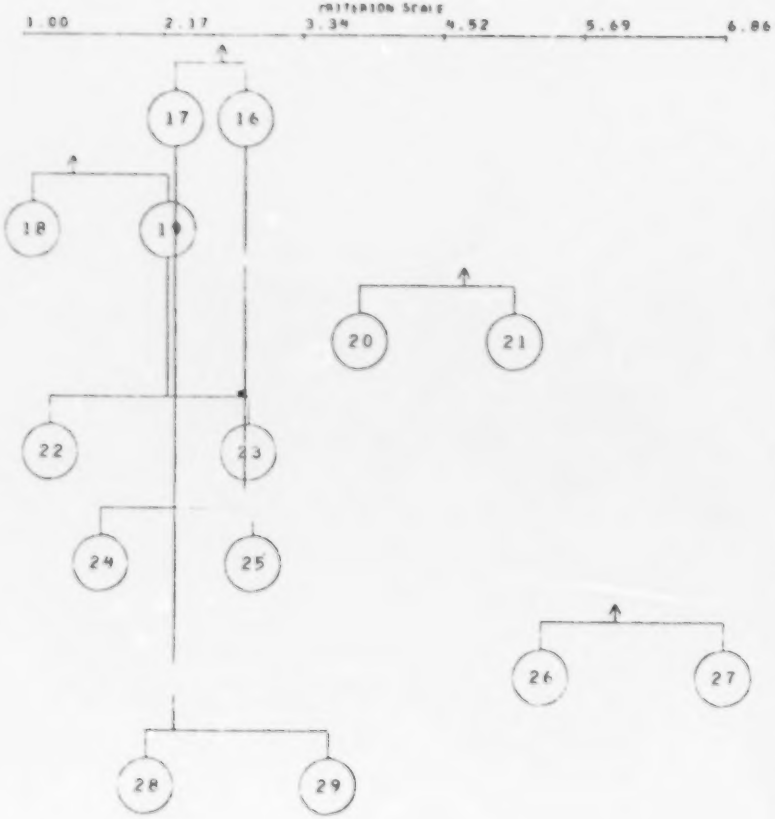
MEAN= 4.21 S.D.= 1.00 N= 87
PREDICTOR VALUES -- 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 34 35 .

PARENT 13 SPLITTING VARIABLE - HUMIDITY

MEAN= 3.68 S.D.= 0.64 N= 41
PREDICTOR VALUES -- 4 14 19 24 25 26 27 28 29 30 31 32 33 34 35 36
FINAL GROUP

MEAN= 4.69 S.D.= 1.02 N= 46
PREDICTOR VALUES -- 37 38 39 40 41 42 43 44 45 46 47 49 51 52 54 62

FIGURE A-1.



SUMMARY CONTINUED	
PARENT 10 SPLITTING VARIABLE - WIND SP	
MEAN= 2.27 S.D.= 0.88 N= 40 PREDICTOR VALUES -- 16 17 18 19 20 21 22 24 25 26 32 33 34 37 38 39 40 43 44 45 63	MEAN= 2.84 S.D.= 0.67 N= 81 PREDICTOR VALUES -- 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
PARENT 4 SPLITTING VARIABLE - HUMIDITY	
MEAN= 1.07 S.D.= 0.34 N= 74 PREDICTOR VALUES -- 0 3 6 8 11 12 13 14 15 16 17 18 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 FINAL GROUP	MEAN= 2.21 S.D.= 1.09 N= 22 PREDICTOR VALUES -- 35 37 38 39 40 41 42 46 47 49 50 51 52 53 54 56 57 61 62
PARENT 15 SPLITTING VARIABLE - TEMP	
MEAN= 3.82 S.D.= 0.94 N= 15 PREDICTOR VALUES -- 19 20 21 24 25 26 27 29 FINAL GROUP	MEAN= 9.11 S.D.= 0.78 N= 11 PREDICTOR VALUES -- 30 31 32 33 34 35 36 37 38 39 40 41 42 44 FINAL GROUP
PARENT 19 SPLITTING VARIABLE - METH	
MEAN= 1.22 S.D.= 0.51 N= 13 PREDICTOR VALUES -- 16 17 18 19 FINAL GROUP	MEAN= 2.09 S.D.= 0.93 N= 17 PREDICTOR VALUES -- 20 21 22 24 27 28 29 30 31 33 34 39 FINAL GROUP
PARENT 16 SPLITTING VARIABLE - THC	
MEAN= 1.65 S.D.= 0.34 N= 5 PREDICTOR VALUES -- 0 1 2 FINAL GROUP	MEAN= 2.93 S.D.= 0.61 N= 76 PREDICTOR VALUES -- 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 30 31 32 a FINAL GROUP
PARENT 7 SPLITTING VARIABLE - CO	
MEAN= 5.34 S.D.= 0.78 N= 16 PREDICTOR VALUES -- 38 39 40 41 42 43 44 45 46 48 FINAL GROUP	MEAN= 4.04 S.D.= 0.92 N= 11 PREDICTOR VALUES -- 31 32 33 35 36 50 54 61 63 FINAL GROUP
PARENT 17 SPLITTING VARIABLE - OZONE	
MEAN= 2.04 S.D.= 0.70 N= 34 PREDICTOR VALUES -- 0 1 2 3 5 6 8 9 11 13 14 15 16 17 22 23 27 29 30 34 39 FINAL GROUP	MEAN= 3.57 S.D.= 0.67 N= 6 PREDICTOR VALUES -- 40 42 44 45 49 FINAL GROUP

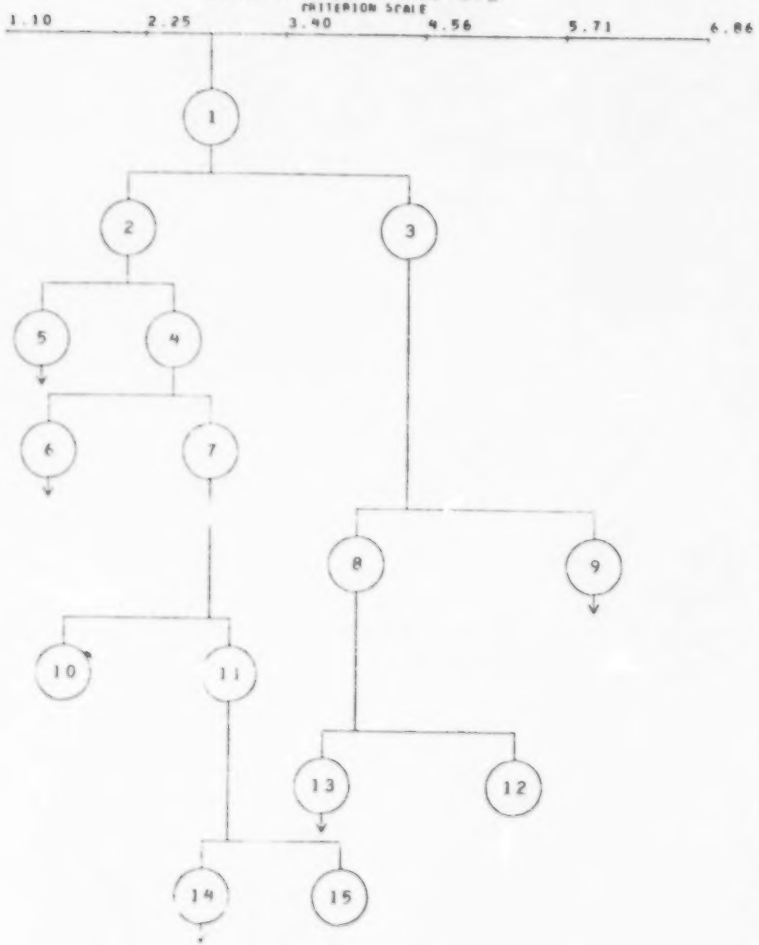
FIGURE A-1. (Continued)

AUTOMATIC INTERACTION DETECTOR

HAZEVECTOR ..... AID ANALYSIS

BINARY TREE STRUCTURE

CRITERION SCALE



SUMMARY TABLE

---TOTAL GROUP---

CRITERION - LIGHT SCAT

TOTAL GROUP N = 432
MEAN = 2.79
STD. DEV. = 1.58

PARENT 1 SPLITTING VARIABLE - NO2

MEAN= 2.11 S.D.= 1.10 N= 305
PREDICTOR VALUES -- 0 1 2 3 4
5 6 7 8 9 10 11 12 13 14 15

MEAN= 4.42 S.D.= 1.36 N= 127
PREDICTOR VALUES -- 16 17 18 19 20
21 22 23 24 25 26 27 28 29 31 32 33
34 35 37 38 39 40 41 42 43 44 47 .

PARENT 2 SPLITTING VARIABLE - WIND VEL

MEAN= 1.90 S.D.= 0.62 N= 106
PREDICTOR VALUES -- 2 3 4 5 6
10 18 19 27 30 31 34 35 37 41 42 43
44 45 50 59 60 61

MEAN= 2.99 S.D.= 1.11 N= 199
PREDICTOR VALUES -- 0 1 8 9 16
17 24 25 26 28 29 32 33 36 40 44 46
50 51 56 57 .

PARENT 4 SPLITTING VARIABLE - HUMIDITY

MEAN= 1.96 S.D.= 0.60 N= 96
PREDICTOR VALUES -- 2 6 7 10 11
12 13 14 15 17 18 20 21 22 23

MEAN= 2.90 S.D.= 1.04 N= 103
PREDICTOR VALUES -- 24 25 26 27 28
29 30 31 32 33 34 35 36 37 38 39 40
41 42 45 46 47 48 49 50 51 52 .

PARENT 5 SPLITTING VARIABLE -

MEAN= 4.01 S.D.= 1.11 N= 100
PREDICTOR VALUES -- 2 3 4 6 8
9 10 11 12 13 14 15 16 17 18 19 20
21 22 23 24 25 26 27 28 29 30 31 .

MEAN= 4.99 S.D.= 1.04 N= 27
PREDICTOR VALUES -- 38 39 40 41 42
43 44 45 46 48 50 52 54 55 58 59 60
61 63

PARENT 7 SPLITTING VARIABLE - TEMP

MEAN= 1.60 S.D.= 0.90 N= 19
PREDICTOR VALUES -- 9 10 12 13 14
15 16

FINAL GROUP

MEAN= 2.97 S.D.= 0.99 N= 134
PREDICTOR VALUES -- 17 18 19 20 21
22 23 24 25 26 27 28 29 30 31 32 33
34 35 36 37 38 39 40 41 42 43 44 .

PARENT 8 SPLITTING VARIABLE - WIND VEL

MEAN= 3.74 S.D.= 0.94 N= 83
PREDICTOR VALUES -- 1 8 9 16 25
27 32 33 34 40 41 48 57 58

MEAN= 6.51 N= 17
PREDICTOR VALUES -- 0 17 24 44 56

FINAL GROUP

PARENT 11 SPLITTING VARIABLE - METH

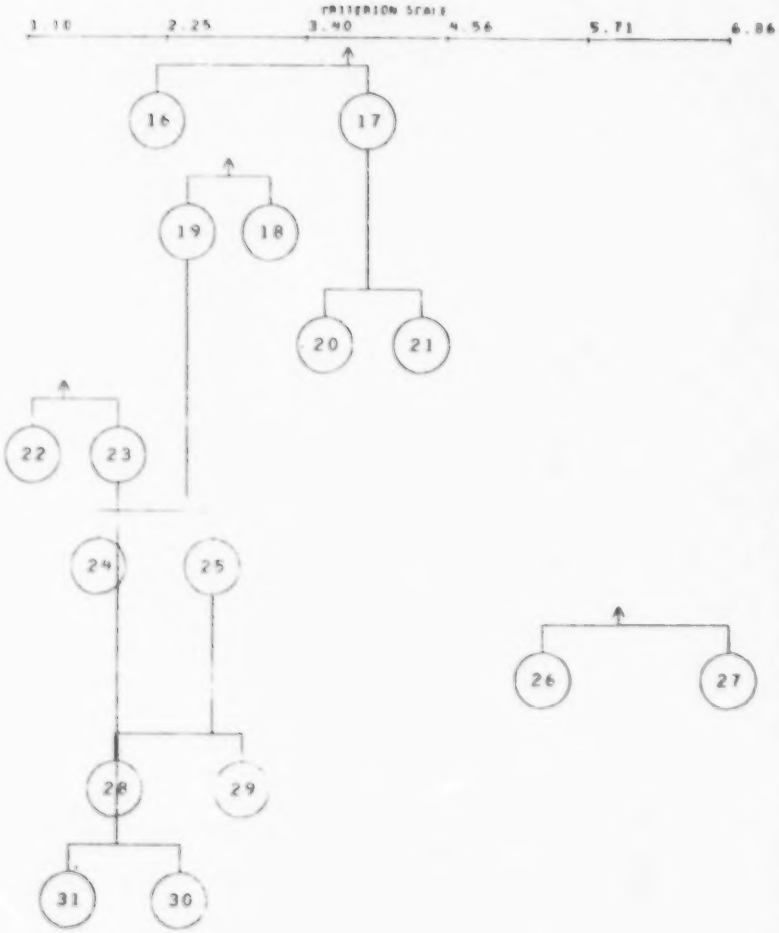
MEAN= 2.76 S.D.= 0.78 N= 189
PREDICTOR VALUES -- 0 1 2 3 4
7 8 7 8 9 10 12 14 15 16 17 18
19 20 21 22 23 24

MEAN= 3.90 S.D.= 1.01 N= 25
PREDICTOR VALUES -- 25 26 27 28 29
30 31 33 34 36 37 46

FINAL GROUP

FIGURE A-2.

HAZEVECTOR ..... AID ANALYSIS
...CONTINUED...

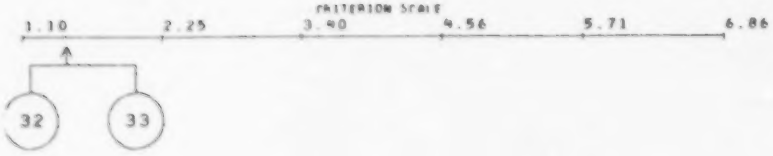


SUMMARY CONTINUED

PARENT 13 SPLITTING VARIABLE - CO	
MEAN= 2.17 S.D.= 0.61 N= 8 PREDICTOR VALUES -- 2 3 4 6 8 9 10 FINAL GROUP	MEAN= 3.90 S.D.= 0.81 N= 94 PREDICTOR VALUES -- 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 34 35 36 FINAL GROUP
PARENT 14 SPLITTING VARIABLE - WIND VEL	
MEAN= 2.92 S.D.= 0.81 N= 54 PREDICTOR VALUES -- 24 25 26 28 32 33 36 40 50	MEAN= 3.10 S.D.= 0.59 N= 55 PREDICTOR VALUES -- 0 1 2 9 14 17 29 40 49 54 57 FINAL GROUP
PARENT 17 SPLITTING VARIABLE - CO	
MEAN= 3.55 S.D.= 0.65 N= 41 PREDICTOR VALUES -- 11 12 13 14 15 16 17 18 19 20 21 22 FINAL GROUP	MEAN= 4.34 S.D.= 0.76 N= 34 PREDICTOR VALUES -- 23 24 25 26 27 28 29 30 31 32 34 35 36 FINAL GROUP
PARENT 5 SPLITTING VARIABLE - TEMP	
MEAN= 1.14 S.D.= 0.39 N= 67 PREDICTOR VALUES -- 4 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 26 27 FINAL GROUP	MEAN= 1.85 S.D.= 0.67 N= 13 PREDICTOR VALUES -- 30 31 32 34 35 36 37 39 40 41 42 43 44 45 46 47 48 49 52
PARENT 19 SPLITTING VARIABLE	
MEAN= 1.69 S.D.= 0.64 N= 12 PREDICTOR VALUES -- 0 4 11 12 14 15 FINAL GROUP	MEAN= 2.63 S.D.= 0.73 N= 42 PREDICTOR VALUES -- 16 18 19 20 21 22 23 24 25 27 28 29 30 32 33 38
PARENT 9 SPLITTING VARIABLE - CO	
MEAN= 4.34 S.D.= 0.78 N= 16 PREDICTOR VALUES -- 38 39 40 41 42 43 44 45 46 48 FINAL GROUP	MEAN= 6.06 S.D.= 0.82 N= 11 PREDICTOR VALUES -- 48 52 54 55 56 58 59 61 63 FINAL GROUP
PARENT 25 SPLITTING VARIABLE - METH	
MEAN= 1.83 S.D.= 0.54 N= 10 PREDICTOR VALUES -- 0 1 2 3 FINAL GROUP	MEAN= 2.88 S.D.= 0.53 N= 24 PREDICTOR VALUES -- 4 5 6 7 8 9 12 14 17 18 19 20 21 22 24 FINAL GROUP
PARENT 23 SPLITTING VARIABLE - WIND VEL	
MEAN= 1.45 S.D.= 0.41 N= 22 PREDICTOR VALUES -- 2 3 4 6 10 37 45 59 60 61 FINAL GROUP	MEAN= 2.37 S.D.= 0.56 N= 17 PREDICTOR VALUES -- 27 34 39 41 42 43 58 FINAL GROUP

FIGURE A-2. (Continued)

HAZEVECTOR ..... AID ANALYSIS
...CONTINUED...



SUMMARY CONTINUED	
PARENT 6 SPLITTING VARIABLE - TEMP	
MEAN= 1.16 S.D.= 0.39 N= 30 PREDICTOR VALUES -- 0 1 2 4 5 6 7 8 9 10 11 12 13 16 17 18 22 23 25 26 28 30 31 33 FINAL GROUP	MEAN= 2.03 S.D.= 0.52 N= 16 PREDICTOR VALUES -- 34 37 41 42 43 44 45 46 51 53 55 FINAL GROUP

FIGURE A-2. (Continued)

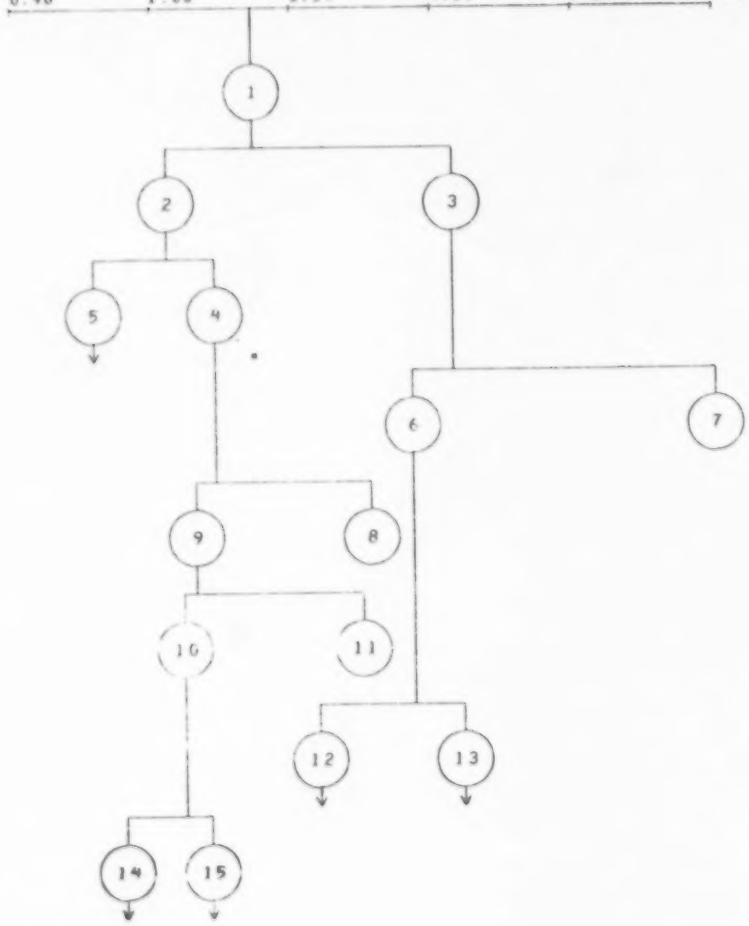
AUTOMATIC INTERACTION DETECTOR

HAZEVECTOR ..... AID ANALYSIS

BINARY TREE STRUCTURE

CRITERION SCALE

0.90 2.00 3.09 4.19 5.28 6.38



SUMMARY TABLE

---TOTAL GROUP---

CRITERION - LIGHT SCAT

TOTAL GROUP N = 432
MEAN = 2.79
STD. DEV. = 1.58

PARENT 1 SPLITTING VARIABLE - THC

MEAN= 2.11 S.D.= 1.11 N= 300
PREDICTOR VALUES -- 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

MEAN= 4.34 S.D.= 1.38 N= 132
PREDICTOR VALUES -- 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 *

PARENT 2 SPLITTING VARIABLE - WIND VEL

MEAN= 1.53 S.D.= 0.77 N= 117
PREDICTOR VALUES -- 2 3 4 5 6 10 18 19 27 28 30 31 34 35 37 41 42 43 44 45 58 59 60 61

MEAN= 2.48 S.D.= 1.14 N= 183
PREDICTOR VALUES -- 0 1 7 8 9 24 25 26 29 32 33 36 40 48 54 55 56 57

PARENT 3 SPLITTING VARIABLE - ACY

MEAN= 4.02 S.D.= 1.14 N= 114
PREDICTOR VALUES -- 0 3 7 10 14 17 21 24 28 31 35

MEAN= 4.38 S.D.= 1.51 N= 18
PREDICTOR VALUES -- 38 42 44 52 53 63

FINAL GROUP

PARENT 4 SPLITTING VARIABLE - WIND VEL

MEAN= 2.32 S.D.= 1.03 N= 162
PREDICTOR VALUES -- 1 8 9 24 25 26 32 33 36 40 48 50 51 56 57

MEAN= 3.69 S.D.= 1.28 N= 21
PREDICTOR VALUES -- 0 10 29 44

FINAL GROUP

PARENT 9 SPLITTING VARIABLE - OZONE

MEAN= 2.22 S.D.= 0.98 N= 150
PREDICTOR VALUES -- 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 25 26 27 29 *

MEAN= 3.42 S.D.= 0.54 N= 12
PREDICTOR VALUES -- 32 33 34 37 38 39 43 44

FINAL GROUP

PARENT 6 SPLITTING VARIABLE - NOX

MEAN= 3.27 S.D.= 0.87 N= 38
PREDICTOR VALUES -- 6 7 8 9 10 11 12 13 14 15 16

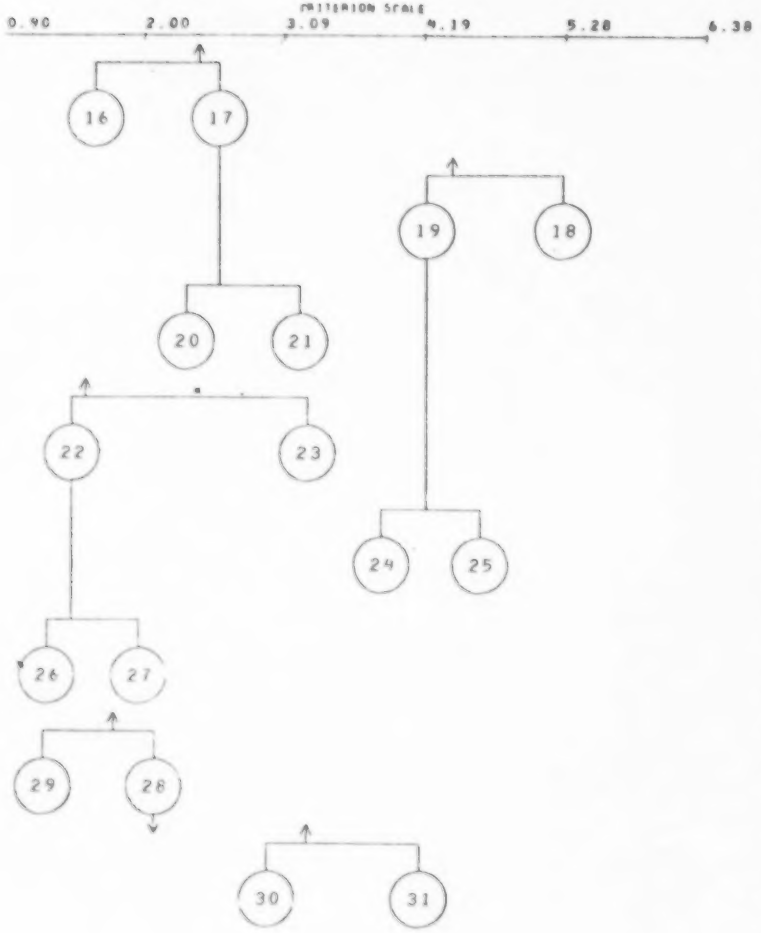
MEAN= 4.40 S.D.= 1.07 N= 76
PREDICTOR VALUES -- 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 37 38 41 42 44 48 52 55 57

PARENT 10 SPLITTING VARIABLE - THC

MEAN= 1.75 S.D.= 0.81 N= 96
PREDICTOR VALUES -- 0 1 2 4 5 6 7 8 9 10 11

MEAN= 2.42 S.D.= 0.90 N= 104
PREDICTOR VALUES -- 12 13 14 15 16 17 18 19 20 21 22 23 24 25

FIGURE A-3.



SUMMARY CONTINUED	
PARENT 15 SPLITTING VARIABLE - OZONE	
MEAN= 1.61 S.D.= 0.88 N= 17 PREDICTOR VALUES -- 0 FINAL GROUP	MEAN= 2.98 S.D.= 0.92 N= 87 PREDICTOR VALUES -- 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 20 21 22 23 25 26 27 29 31
PARENT 13 SPLITTING VARIABLE - WIND VEL	
MEAN= 4.20 S.D.= 1.00 N= 62 PREDICTOR VALUES -- 0 8 25 32 33 40 41 48 57	MEAN= 5.26 S.D.= 0.92 N= 14 PREDICTOR VALUES -- 16 24 49 56 FINAL GROUP
PARENT 17 SPLITTING VARIABLE - NOX	
MEAN= 2.32 S.D.= 0.85 N= 61 PREDICTOR VALUES -- 4 5 6 7 8 9 10 11 12 13 14 15 16 FINAL GROUP	MEAN= 3.21 S.D.= 0.75 N= 26 PREDICTOR VALUES -- 17 18 22 23 24 25 26 27 28 29 30 36 37 42 48 FINAL GROUP
PARENT 5 SPLITTING VARIABLE - OZONE	
MEAN= 1.42 S.D.= 0.63 N= 110 PREDICTOR VALUES -- 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 26 29 30 FINAL GROUP	MEAN= 3.27 S.D.= 0.69 N= 7 PREDICTOR VALUES -- 38 39 40 42 44 46 FINAL GROUP
PARENT 19 SPLITTING VARIABLE - OZONE	
MEAN= 3.05 S.D.= 1.05 N= 34 PREDICTOR VALUES -- 0 FINAL GROUP	MEAN= 4.62 S.D.= 0.75 N= 26 PREDICTOR VALUES -- 1 2 3 4 5 6 9 10 11 14 21 25 28 30 31 34 45 63 FINAL GROUP
PARENT 22 SPLITTING VARIABLE - SO2	
MEAN= 1.23 S.D.= 0.52 N= 81 PREDICTOR VALUES -- 0 1 3 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 FINAL GROUP	MEAN= 1.95 S.D.= 0.59 N= 25 PREDICTOR VALUES -- 25 26 27 28 29 30 31 32 38 40 41 FINAL GROUP
PARENT 14 SPLITTING VARIABLE - WIND VEL	
MEAN= 1.20 S.D.= 0.30 N= 17 PREDICTOR VALUES -- 25 32 40 50 56 FINAL GROUP	MEAN= 2.07 S.D.= 0.84 N= 29 PREDICTOR VALUES -- 1 8 9 24 26 33 48 57
PARENT 12 SPLITTING VARIABLE - METH	
MEAN= 2.96 S.D.= 0.97 N= 28 PREDICTOR VALUES -- 5 6 7 8 9 10 12 14 15 16 17 18 19 20 21 24 FINAL GROUP	MEAN= 4.15 S.D.= 0.94 N= 10 PREDICTOR VALUES -- 25 26 27 28 30 32 33 37 46 FINAL GROUP

FIGURE A-3. (Continued)

HAZEVECTOR ..... AID ANALYSIS
...CONTINUED...

SUMMARY CONTINUED	
PARENT 28 SPLITTING VARIABLE - NO	
MEAN= 0.98 S.D.= 0.17 N= 5 PREDICTOR VALUES -- 1 2 3 FINAL GROUP	MEAN= 2.30 S.D.= 0.74 N= 24 PREDICTOR VALUES -- 4 5 6 7 8 9 11 12 15 16 17 31 FINAL GROUP

FIGURE A-3. (Continued)

and a standard deviation of 1.10; the corresponding information for group 3 is 127 observations, a mean of 4.42, and a standard deviation of 1.36. Further, the table lists the values of the predictor variable (NO_2) associated with subgroups 2 and 3. The coded values 0 through 15 are associated with group 2, and the coded values from 16 through 63 are associated with group 3. Asterisks at the end of a list signify that more coded values follow, but no space exists to complete the list. Each predictor variable is coded initially into integer values in the range 0 to 63. The coding is done as follows: let R denote the range of the observed values, let z denote the minimum of the observed values, and let x denote a predictor value; the coded value c is given by

$$c = \left[62.5 \frac{x - z}{R} + 0.5 \right] ,$$

where $[\alpha]$ denotes the greatest integer less than or equal to α.

In words, one may describe the first split loosely as follows: when NO_2 values are in the lower quartile of their range, the corresponding light-scattering readings have a mean of 2.11; when the NO_2 values are higher, the corresponding light-scattering readings have a mean of 4.42.

The remaining splits may be interpreted in a similar fashion. The classification rules in the tree may be verbalized for parts of the tree. For example, one can say that the average light scattering is 5.96 given that the NO_2 reading is in the upper 75 percent of the NO_2 range and the CO reading is in the upper 37 percent of the CO range. As a practical way of choosing important variables and interactions from an AID tree, one generally observes the variables associated with the first "several" splits. It is possible by utilizing various stopping rules, formal statistical tests, and iterative procedures to refine an AID tree to a well-formed rule. However, the purpose here was to select several variables for the time-series analysis which would explain a portion of the variance in the light scattering, so the initial trees suffice.

Although temperature shows up as a potentially important splitter in this AID analysis, subsequent attempts to develop a transfer function relating light scattering and temperature failed to obtain a significant relationship. A possible explanation may be that temperature is only significant in its interaction with NO_2 in the analysis, and no attempt was made to construct interaction variables for the time-series analysis.

The wind-speed (WIND SP) variable in Figure A-1 was obtained by first computing a wind-velocity vector. The 10-minute wind-speed and wind-direction values were formed into vectors, and an hourly average wind-velocity vector was computed using vector addition. The wind-speed variable used in the analysis was simply the wind-speed component of this vector. A previous analysis had used a straight averaged wind speed. However, neither the previous analysis nor this one showed much significance for wind speed.

Figure A-2 shows the results of an AID analysis of the sample when the wind-velocity vector (WIND VEL) replaces wind speed. The coding scheme for wind velocity is necessarily different from that used for the other variables. Since a wind-velocity value may be associated with a point in the plane, a portion of the plane including all the wind-velocity values in the data was divided into segments. This procedure is illustrated in Figure A-4. The plane is in polar coordinates, with the r-coordinate representing the wind-speed component, the θ-coordinate representing the wind-direction component, and 0° and 360° representing due north. Concentric circles are drawn in increments of 2 mph to a maximum of 16 mph. The circles are divided into octants and the resulting partition of this portion of the plane coded as shown. As an example,

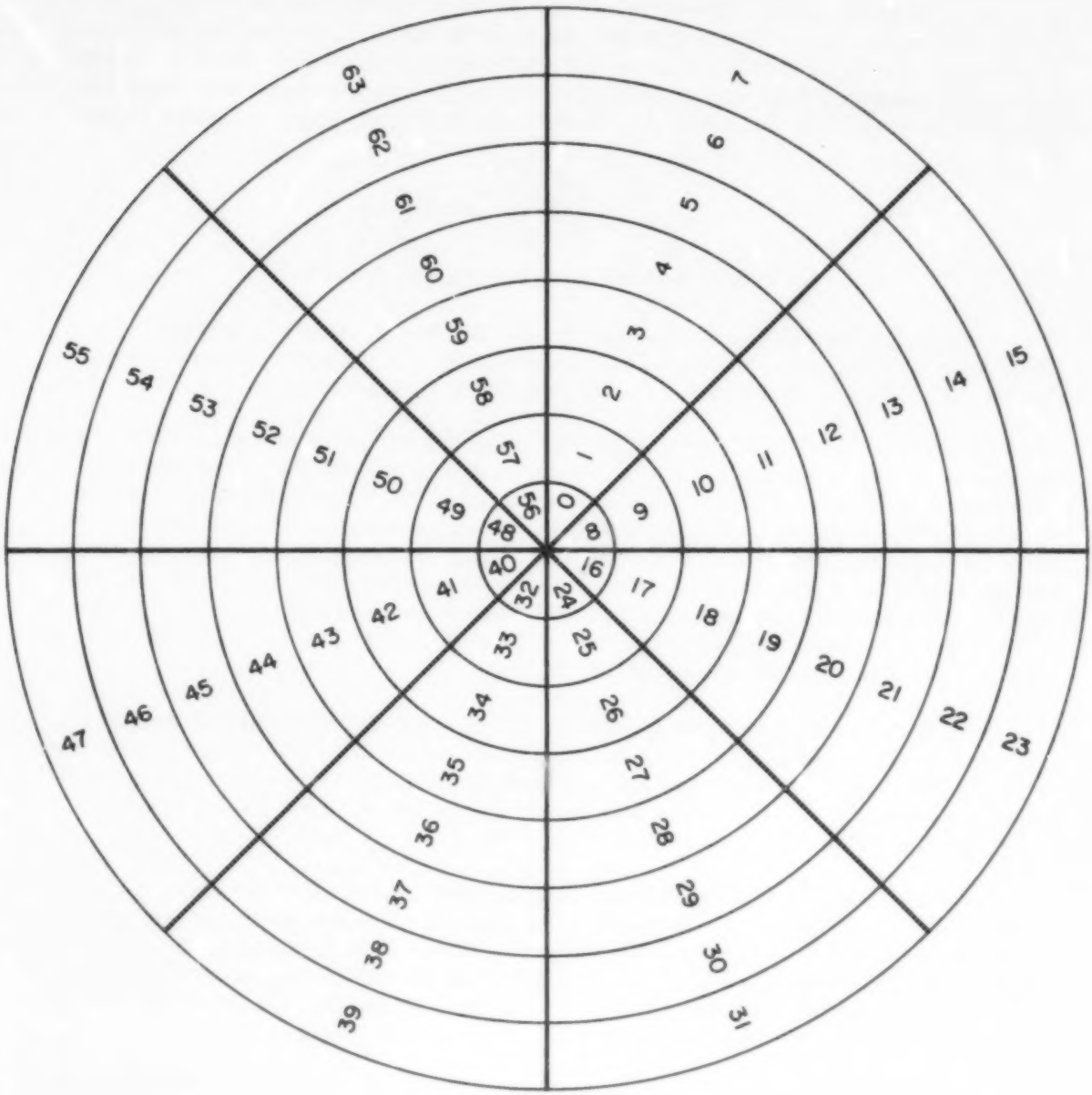


FIGURE A-4. CODING SCHEME USED TO DEVELOP WIND DIRECTION AND VELOCITY VECTORS FOR SUBSEQUENT ANALYSIS OF WIND EFFECTS ON LIGHT SCATTERING

a wind-velocity vector of (4.5, 100°) is coded as 18. Since no order relationship exists in this code, potential splits on wind velocity could occur on an arbitrary division of wind-velocity codes into two subsets, whereas splits on other variables were restricted to being monotone [i.e., all values of the predictor associated with one subgroup of a split must be greater than (or less than) all values associated with the other subgroup.]

Figure A-2 shows the increased power of the new variable. Wind velocity becomes the second most powerful splitter and also occurs elsewhere in the tree. Figure A-5 illustrates the split of group 2 into groups 4 and 5 on wind velocity. The hatched areas are associated with group 5 and the dotted areas are associated with group 4. This figure illustrates the tendency of light scattering to be higher when the wind-speed component is lower, but also shows that wind direction is a factor in separating "lower" from "higher" wind speeds.

While this analysis is highly informative, and shows the power which certain interactions may have, it was not possible to include the vector-valued wind-velocity variable in the time-series analysis. Devising a method to incorporate such variables is left to future studies.

An effect called "masking" can sometimes occur in AID analyses. This occurs when one variable is correlated with other "stronger" variables in the AID sense. While such a variable may be useful in time-series analysis it may not show up initially as a strong variable in an AID analysis. One such variable in this data is THC, and this is shown in Figure A-3. This run was made after removing the variables NO_2, CO, RH, and temperature from the set of predictor variables.

As a result of the AID runs and some preliminary time-series analysis, it was decided to use NO_2, CO, RH, and THC as predictor variables in a transfer-function development for light scattering.

Time-Series Analysis

A time series is a sequence of observations, denoted $\{z_t\}_{t=1}^{N}$, taken at equally spaced time intervals. The mathematical formulation of the Box-Jenkins univariate time-series model is given by

$$\Phi_P(B^s)\phi_p(B)\nabla_s^D\nabla^d z_t = \Theta_0 + \Theta_Q(B^s)\theta_q(B)a_t \qquad (1)$$

where

s denotes the length of the period of seasonality;

B is the backward shift operator defined by $Bz_t = z_{t-1}$;

∇, B^s, ∇_s, ∇^d, ∇_s^D are operators defined in terms of B by

$\nabla = 1 - B$ (backward difference operator),

$B^s = B(B^{s-1})$ for $s \geq 1$ where $B^0 \equiv 1$,

$\nabla_s = 1 - B_s$ (backward seasonal difference operator),

$\nabla^d = \nabla(\nabla^{d-1})$ for $d \geq 1$ where $\nabla^0 \equiv 1$,

$\nabla_s^D = \nabla_s(\nabla_s^{D-1})$ for $D \geq 1$ where $\nabla_s^0 \equiv 1$;

Φ_P, ϕ_p, Θ_Q, θ_q are polynomials with constant terms equal unity of degrees P, p, Q, and q in their respective arguments;

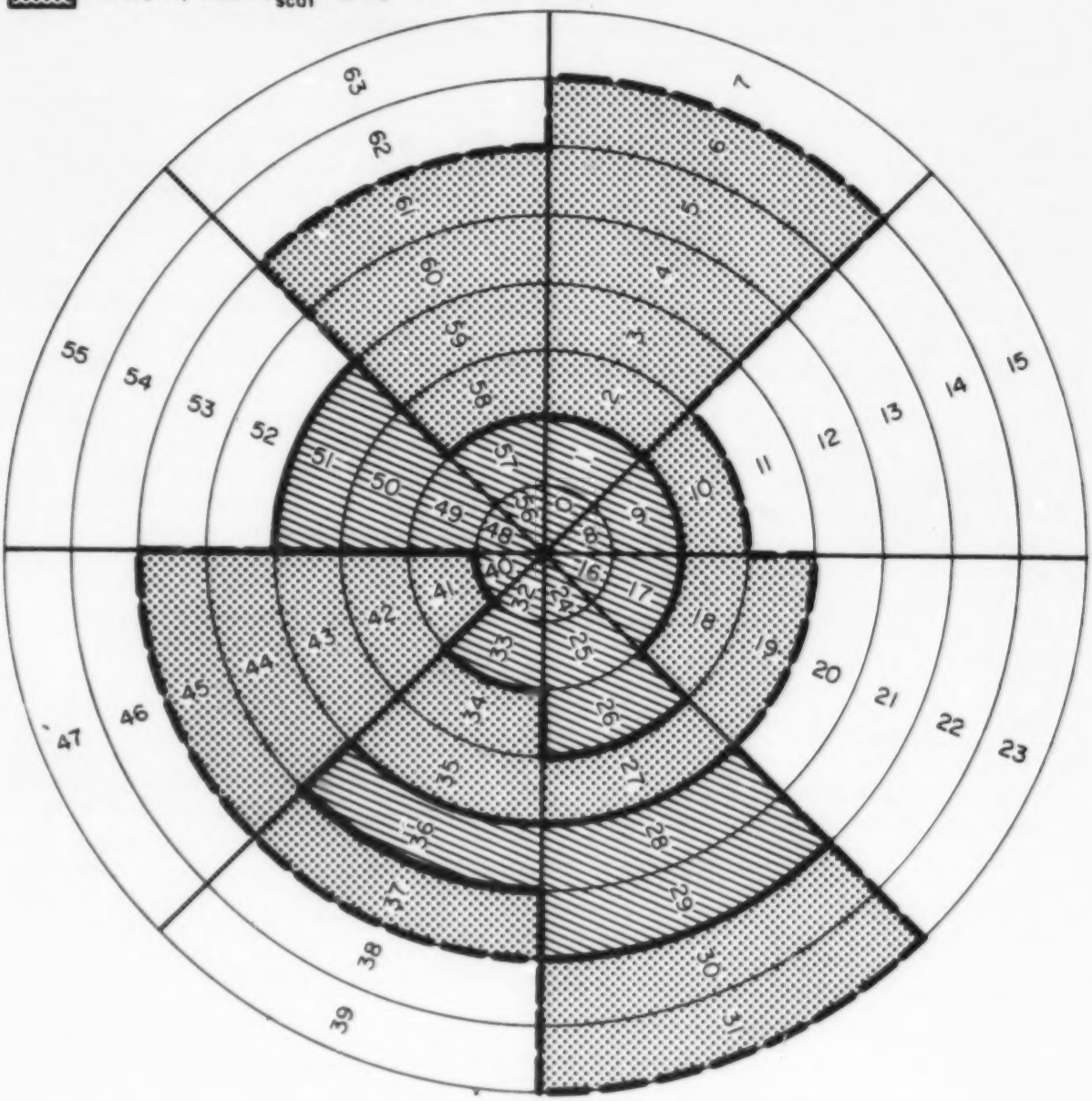


FIGURE A-5. SCHEMATIC REPRESENTATION OF AID RESULTS WHERE WIND VELOCITY IS A DIRECTIONAL VELOCITY VECTOR

Θ_0 is a constant;

a_t is a sequence of "white noise", that is, a_t is $N(0,\sigma_a)$ and is uncorrelated for different times.

To illustrate the meaning of these operators and polynomials, we give some examples:

(a) $\nabla z_t = (1 - B)z_t = z_t - z_{t-1}$

(b) $B^{24}z_t = z_{t-24}$

(c) $\nabla_{24} z_t = z_t - z_{t-24}$

(d) $\nabla^2 z_t = \nabla(z_t - z_{t-1})$

$= z_t - z_{t-1} - (z_{t-1} - z_{t-2})$

$= z_t - 2z_{t-1} + z_{t-2}$

(e) $\phi_2(B) = (1 - .75B + .5B^2)$

$\phi_2(B)z_t = z_t - .75z_{t-1} + .5_{t-2}$.

A shorthand notation for (1) if $\Theta_0 = 0$ is (p, d, q) X (P, D, Q)$_s$. $\Theta_0 \neq 0$ signifies a deterministic trend. $\phi_p(B)$ and $\Phi_P(B^s)$ are called autoregressive operators; $\theta_q(B)$ and $\Theta_Q(B^s)$ are called moving average operators. Differencing and seasonal differencing may be necessary to induce stationarity in the time series. In most applications, p, d, q, P, D, and Q are less than or equal to 2.

Tables A-1 and A-2 contain the results of fitting Box Jenkins models to the light scattering, NO_2, CO, RH, and THC series. The data used in developing the univariate models were hourly averages from August 22 to August 30, 1972, collected on Welfare Island, New York City. Table A-1 contains the models developed incorporating seasonal differencing; Table A-2 contains models developed without seasonal differencing. Table A-2 was developed in order to see whether the models could be simplified, since Table A-1 shows evidence of "overfitting". As will be seen later, seasonal differencing of the data gives superior results in developing transfer functions.

The univariate analysis performed is preliminary to the multivariate analysis, i.e., the development of transfer-function models. The general form of the transfer-function model is

$$\nabla^d \nabla_s^D y_t = \mu + \sum_{i=1}^{M} \frac{\omega_i(B)\Omega_i(B^s)}{\delta_i(B)\Delta_i(B^s)} \nabla^d \nabla_s^D x_{i,t-b_i} + \frac{\theta(B)\Theta(B^s)}{\phi(B)\Phi(B)^s} a_t$$

where

$\{y_t\}$ designates the output (dependent) time series;

$\{x_{i,t}\}$ designates the i^{th} input (independent, predictor) time series;

b_i designates the time lag between input series i and the output series;

μ is a constant;

$\omega_i(B)$, $\delta_i(B)$, $\Omega_i(B^s)$, and $\Delta_i(B^s)$ are polynomials in their respective arguments (usually of degree ≤ 2);

all other symbols are as previously defined.

TABLE A-1. SUMMARY OF UNIVARIATE ANALYSIS (WITH SEASONAL DIFFERENCING)

	Model	Parameters	Residual Variance	Residual Autocorrelation χ^2/Degrees of Freedom
Light Scattering	$(2, 1, 0)\ (0, 1, 2)_{24}$	$\phi_1 = .203 \pm .07^*$ $\phi_2 = -.338 \pm .07$ $\Theta_1 = 1.063 \pm .07$ $\Theta_2 = -.182 \pm .06$	.282	83.09/56
NO_2	$(2, 1, 0)\ (0, 1, 1)_{24}$	$\phi_1 = -.131 \pm .07$ $\phi_2 = -.290 \pm .07$ $\Theta_1 = .900 \pm .02$	$.133 \times 10^{-3}$	47.71/57
CO	$(0, 1, 2)\ (0, 1, 1)_{24}$	$\theta_1 = -.255 \pm .07$ $\theta_2 = .237 \pm .07$ $\Theta_1 = .896 \pm .02$	.116	47.57/57
RH	$(2, 1, 0)\ (0, 1, 1)_{24}$	$\phi_1 = .115 \pm .07$ $\phi_2 = -.130 \pm .07$ $\Theta_1 = .852 \pm .03$	11.196	40.57/57
THC	$(0, 1, 0)\ (0, 1, 1)_{24}$	$\Theta_1 = .862 \pm .025$	.120	47.27/59

* ± standard deviation of parameter estimate.

TABLE A-2. SUMMARY OF UNIVARIATE ANALYSIS (WITHOUT SEASONAL DIFFERENCING)

	Model	Parameters	Residual Variance	Residual Autocorrelation χ^2/Degrees of Freedom
Light Scattering	$(2, 1, 0)\ (1, 0, 0)_{24}$	$\phi_1 = .211 \pm .07^*$ $\phi_2 = -.244 \pm .07$ $\Phi_1 = .050 \pm .066$	.300	83.67/57
NO_2	$(2, 1, 0)\ (1, 0, 0)_{24}$	$\phi_1 = -.094 \pm .07$ $\phi_2 = -.224 \pm .07$ $\Phi_1 = .285 \pm .06$	$.140 \times 10^{-3}$	61.73/57
CO	$(0, 1, 2)\ (1, 0, 0)_{24}$	$\theta_1 = -.288 \pm .07$ $\theta_2 = .168 \pm .07$ $\Phi_1 = .203 \pm .06$	.137	55.28/57
RH	$(0, 1, 1)\ (1, 0, 0)_{24}$	$\theta_1 = -.240 \pm .13$ $\Phi_1 = .132 \pm .07$	12.300	54.08/58
THC	$(0, 1, 0)\ (1, 0, 0)_{24}$	$\Phi_1 = .196 \pm .07$	.129	39.78/59

* ± standard deviation of parameter estimate.

The coefficients of the polynomials $\omega_i(B)$, $\Omega_i(B^s)$, $\delta_i(B)$, $\Delta_i(B^s)$ are determined from the cross correlations of various lags between the prewhitened input series and the transformed output series. The prewhitened input series are the residual series of the corresponding univariate time-series models for the input series. The transformed output series is obtained by operating on the output series with the same formal model for each input series. The noise series is then modeled by univariate analysis and all parameters refit by a least squares analysis.

The χ^2/Degrees of Freedom values listed in Tables A-1 through A-4 are obtained from a "portmanteau" test of the hypothesis of model adequacy. This test is described by Box and Jenkins[32]. Some of the values obtained may seem higher than ones generated from white noise, but an examination of the data shows that one or two large values obtained from differencing can have a severe impact on the "portmanteau" χ^2, and that trying to reduce the χ^2 statistic by further fitting would only distort the model.

Tables A-3 and A-4 contain the statistical information on the transfer-function models developed. The models in Table A-3 were developed with seasonal differencing; the models in Table A-4 were not seasonal differenced.

TABLE A-3. SUMMARY OF TRANSFER FUNCTION ANALYSIS (WITH SEASONAL DIFFERENCING)

		Transfer Function			Noise Model			Cross Correlation χ^2/DF (X_i-Noise)
		ω_0	ω_1	ω_2	θ_1	θ_2	Θ_1	
I	X_1 = NO_2	12.670 ± 2.9*	-10.981 ± 2.8	0.	-.153 ± .08	.127 ± .08	.670 ± .05	85.15/48
	X_2 = CO	.653 ± .09	.179 ± .09	.232 ± .08	Residual Variance = .146			70.25/47
	X_3 = RH	.042 ± .009	0.	0.	χ^2/DF (Noise-Noise) = 50.92/48			
II	X_1 = NO_2	12.434 ± 3.0	-8.104 ± 3.0	0.	-.256 ± .08	.074 ± .08	.711 ± .05	80.83/48
	X_2 = CO	.573 ± .10	.234 ± .11	.291 ± .09	Residual Variance = .155			70.78/47
	X_3 = THC	.273 ± .10	-.253 ± .1	.030 ± .10	χ^2/DF (Noise-Noise) = 49.03/48			55.69/47
III	X_1 = NO_2	19.740 ± 2.9	-6.904 ± 2.9	0.	-.027 ± .09	.147 ± .08	.681 ± .06	84.33/48
	X_2 = RH	.053 ± .01	0.	0.	Residual Variance = .182			45.41/49
	X_3 = THC	.303 ± .1	-.255 ± .1	0.	χ^2/DF (Noise-Noise) = 53.19/48			56.30/47
IV	X_1 = CO	.719 ± .09	.005 ± .09	.275 ± .09	.251 ± .08	.118 ± .08	.671 ± .05	66.24/47
	X_2 = RH	.037 ± .009	0.	0.	Residual Variance = .157			46.22/49
	X_3 = THC	.272 ± .1	-.225 ± .1	.088 ± .1	χ^2/DF (Noise-Noise) = 40.60/48			52.36/47

* ± standard deviation of parameter estimate.

TABLE A-4. SUMMARY OF TRANSFER FUNCTION ANALYSIS (WITHOUT SEASONAL DIFFERENCING)

		Transfer Function			Noise Model		Cross Correlation χ^2/DF (X_i-Noise)
		ω_1	ω_1	ω_2	θ_1	θ_2	
I	X_1 = NO_2	14.239 ± 2.72*	0.	0.	-.072 ± .07	.101 ± .07	68.28/49
	X_2 = CO	.468 ± .09	.032 ± .08	.129 ± .07	Residual Variance = .155		57.34/47
	X_3 = RH	.046 ± .01	-.018 ± .01	0.	χ^2/DF (Noise-Noise) = 68.30/48		37.01/48
II	X_1 = NO_2	12.122 ± 2.9	0.	0.	-.123 ± 0.7	.100 ± .07	53.45/49
	X_2 = CO	.456 ± .09	.089 ± .08	.210 ± .08	Residual Variance = .169		55.10/47
	X_3 = THC	.346 ± .09	-.239 ± .09	0.	χ^2/DF (Noise-Noise) = 82.57/48		48.61/48
III	X_1 = NO_2	18.895 ± 2.6	0.	0.	-.033 ± .07	.122 ± .07	63.98/49
	X_2 = RH	.042 ± .01	-.018 ± .01	0.	Residual Variance = .172		39.52/48
	X_3 = THC	.251 ± .09	-.169 ± .09	0.	χ^2/DF (Noise-Noise) = 79.35/48		47.45/48
IV	X_1 = CO	.614 ± .08	-.002 ± .08	.142 ± .08	-.109 ± .07	.136 ± .07	60/87/47
	X_2 = RH	.042 ± .01	-.004 ± .01	0.	Residual Variance = .165		33.80/48
	X_3 = THC	.312 ± .09	-.137 ± .09	0.	χ^2/DF (Noise-Noise) = 75.49/48		39.94/48

* ± standard deviation of parameter estimate.

APPENDIX B

ORGANIC ANALYTICAL PROCEDURES

Solvent Extraction of Particulate Matter

Samples were subjected to either methylene chloride soxhlet-extraction (20 hours) or to sequential soxhlet-extraction first using methylene chloride (20 hours) and then dioxane (44 hours). High purity "distilled-in-glass" solvents were obtained from Burdick and Jackson, Muskegon Michigan; dioxane was additionally redistilled before use.

Extractions were conducted in soxhlet apparatus of 50 ml capacity fitted with 100 ml RB flasks; 25 mm x 80 mm thimbles were used. Thimbles were preextracted with methylene chloride (48 hours) and methanol (48 hours) before use. Up to three 6-inch-diameter filter disks or one 8-inch x 10-inch filter could be extracted in one thimble. For each analytical extraction 100 ml of solvent was employed.

The volume of each methylene chloride extract was reduced to approximately 0.5 ml by vacuum distillation (250 mm Hg at 35 C pot temperature). This was accomplished using 25 ml graduated concentrator tubes fitted with Kuderna-Danish columns, obtained from Kontes Glass Company, Vineland, New Jersey. Samples were heated under vacuum using a 10-tube Rotary-Evapomix evaporator obtained from Buchler Instruments, New York, New York; rubber hose connections were replaced with Teflon.

The weight of methylene chloride extractable matter was determined by taking a small aliquot of the concentrate (~5 percent) and evaporating the solvent on a light (~5 mg) aluminum weighing pan. Rapid evaporation of the methylene chloride led to collection of a small quantity of moisture on the aluminum pan and residue. This was permitted to dissipate by storage of the pans overnight in a desicator; the residue weight was then determined. Pans were tared and reweighed using a Cahn electrobalance, and aliquot weights were determined to ±2 μg. The fraction represented by the aliquot was determined gravimetrically. That is, the total weight of the concentrate (solution) was determined. Next, a volume was withdrawn in a microsyringe (i.e., 25 μl). The syringe was then weighed before and after the solution was dispensed onto the weighing pan. Using these data the total weight of the extractable matter was calculated. Use of this procedure obviated the necessity of taking the entire methylene chloride extract to dryness.

Dioxane extracts were concentrated by lyophilization (freeze drying). An all glass apparatus was used, and extracts were lyophilized in tared 25 ml RB flasks. Use of vacuum grease (silicone) was minimized by utilization of Teflon joint tape.

Solvent and filter blanks were carried through the reflux and concentration procedures. In calculating the weight percent solvent extractable, values due to such solvent and filter background were subtracted. Data concerning blanks are presented in the Results and Discussion section. Values for weight percent methylene chloride extractable and weight percent dioxane extractable were calculated as follows:

$$\text{Weight percent solvent extractable} = \frac{\text{(Weight of extracted matter, corrected)}}{\text{(Weight of total particulate)}} \times 100 \ .$$

In order to demonstrate that the dioxane does not extract significant quantities of inorganic salt from the filters, 100-ml portions of dioxane were stirred with ammonium sulfate, ammonium nitrate, lead nitrate, sodium carbonate, and sodium chloride. The suspensions were filtered and the dioxane filtrates were lyophilized in tared flasks. Salt residues were only barely visible and did not exceed 0.3 mg.

Determination of Weight Percent CHN

Determinations of weight percent C, H, and N were conducted using a Perkin-Elmer Model 240 elemental analyzer with gas purification accessory. The instrument performs automated Pregel-Dumas determinations. Values shown for weight percent oxygen [O] were calculated by difference from the CHN data. Approximately 1 mg of sample was used for each determination.

Infrared Spectroscopy

Infrared spectra were obtained using a Perkin-Elmer Model 521 grating infrared spectrophotometer. Spectra were obtained of thin films of sample on a micro sodium chloride plate (13 mm x 4 mm). Typically ~ 1.5 mg of extractable matter was used. Films were deposited from concentrated extract solutions, after which the last traces of solvent were removed under vacuum. Typically samples were maintained in vacuo for 1 hour before spectra were obtained. It was demonstrated that extended maintenance of the sample in vacuo (i.e., 3 hr) had the effect of uniformly reducing the intensities of the IR bands. Thus, it was concluded that removal of solvent traces under vacuum does not fractionate the sample and is an acceptable procedure in this application.

Instrument operating conditions included a slit program setting of 1000 units to produce a varying spectral slit width of 2-6 cm^{-1}. The gain setting was adjusted to give a 1 percent overshoot for a 10 percent pen deflection. The scan rate setting gave a rate of 6 cm^{-1}/sec above 2000 cm^{-1} and 3 cm^{-1}/sec below 2000 cm^{-1}. The attenuation was 3 sec full scale. The total instrument was purged with dry nitrogen to eliminate any lack of compensation between the two beams due to water vapor. Amplifier balance was set for essentially zero drift with both beams blocked.

In practically all cases film thicknesses were adjusted so that the intensity of all measured bands (with the exception of that for CH stretching at 2920 cm^{-1}) fell in the desired range of 30-60 percent transmittance. In some cases it was not possible to obtain spectra having all bands in the desired intensity range. Thus spectra were obtained to give the maximum number of key bands in this desired range.

Reduction of spectroscopic results to numerical form was performed by calculating relative peak intensities for specified absorptions. Values of relative peak intensity were obtained by calculating the ratio of the optical density for the specified absorption to that observed for the CH stretching vibration at 2920 cm^{-1}; i.e.,

$$\text{Relative intensity} = \frac{\text{(O.D. - specified absorption)}}{\text{(O.D. - 2920 cm}^{-1}\text{)}}$$

Nuclear Magnetic Resonance Spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy was performed using a Varian Associates HA 60-IL spectrometer (60 MHz) or a JOEL PS-100 spectrometer (100 MHz). Both instruments were equipped to perform Fourier-Transform (FT) spectroscopy. Instrument control, data acquisition, and reduction in the Fourier-transform mode were accomplished using a Digilab digital computer interfaced with the spectrometers.

Spectra were obtained using the unfractionated methylene chloride extractable matter. A portion of concentrated extract was taken containing ~ 1.5 mg of residue. The solution was taken to dryness in a small conical tube under a gentle stream of dry nitrogen. After the sample was no longer fluid, the tarry residue was maintained under the nitrogen stream for an additional 15-20 min. The sample was then dissolved in 40-50 μl of $CDCl_3$ (Merck, silver-leaf stabilized) and transferred to a 1.8-mm ID capillary tube. The capillary tube was sealed and placed in a standard 5-mm OD x 5-inch NMR tube. Teflon spacers were used to maintain the capillary tube concentric with the larger tube.

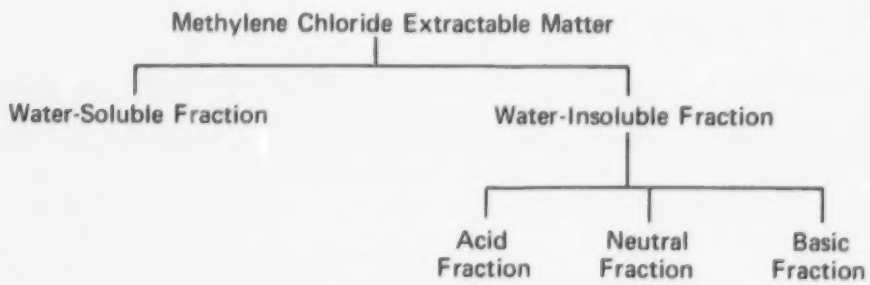
When samples were run on the Varian spectrometer 100 μl of hexafluorobenzene was placed in the outer tube annulus as a source of fluorine resonance for field stabilization. When the JOEL instrument was employed, $CDCl_3$ served as the field stabilization source. Tetramethylsilane was used as the internal standard.

All spectra were obtained in the FT mode. Only the $^1H_1$ nucleus was observed. Typical operational parameters were a 1 μsec pulse width (where a 90-degree flip angle required 20 μsec) with a pulse-to-pulse delay time of 1.5 seconds. The bandwidth was 2000 Hz with 1096 data points collected per pulse, to give a resolution of 0.98 Hz. Typically, computer-time-averaging of 500 pulses was conducted for each sample. The spectral data were integrated both digitally and in the analogue mode, and the integrated values checked to within ±2 percent. From the integrated values, aromatic/aliphatic ratios were obtained. These were calculated on the basis of integrated resonances for methyl and methylene protons and integrated resonances for aromatic protons; i.e.,

$$\text{Aromatic/Aliphatic Ratio} = \frac{\text{(Integrated resonances:aromatic protons)}}{\text{(Integrated resonances:methyl + methylene protons)}}.$$

Sample Fractionation

The methylene chloride extractable matter was fractionated as noted in the scheme below.



Fractionation into water-soluble and water-insoluble components was accomplished by extracting the methylene chloride solution (~3 ml) four times with 2-ml portions of distilled water. All extractions were conducted in conical centrifuge tubes agitated using a Vortex mixer. This procedure permits convenient centrifugation of emulsions prior to separation of phases and minimizes sample losses. Water was removed by lyophilization, and the weight of the water-soluble fraction was determined.

The water-insoluble material remaining in solution was next extracted four times with 2-ml portions of 2N aqueous sodium hydroxide and twice with 2-ml portions of distilled water. The sodium hydroxide and water washes were combined. The methylene chloride solution was then extracted three times with 2-ml portions of 2N hydrochloric acid and twice with 2-ml portions of distilled water. The acid and water washes were combined. Additional water washes were made until neutral; these were discarded. Material remaining in methylene chloride solution after extraction with both aqueous sodium hydroxide and hydrochloric acid is defined as the water-insoluble neutral fraction. Although not conducted during this program, it is recommended that the above acid and base aqueous extracts be backwashed with methylene chloride before neutralization to prevent any slight carry-over of neutral products to the acid and/or basic fractions.

The sodium hydroxide extract (containing organic-acid salts) was brought to pH 0.1, and the free acid was extracted into methylene chloride using a continuous liquid/liquid extractor for 96 hours. Similarly, the hydrochloric acid extract (containing organic-base salts) was brought to pH 13, and the free base was extracted into methylene chloride, again using a continuous liquid/liquid extractor for 96 hours. The liquid/liquid extractor employed was designed for use with solvents denser than water. It is similar to the commercially available Hershberg/Wolfe extraction apparatus except that the apparatus employed is of smaller capacity (i.e., 70 ml methylene chloride and 20 ml aqueous phase) and in place of a sintered glass frit, the top of the aqueous column was agitated using a small magnetic stirring bar (i.e., the drive magnet was positioned at the side of the apparatus).

In order to determine the acid/base/neutral distribution, the methylene chloride solutions of these fractions were concentrated as described above, aliquot weights were determined, and the total weight of each fraction was calculated. In preparation for alcohol determination the methylene chloride solution of the neutral fraction was dried before the final concentration step. Drying was accomplished by refluxing the solution over a 3-A molecular sieve contained in the glass thimble of a small soxhlet apparatus (i.e., "micro" or Bantamware size, 30-ml flask capacity). Drying the solution directly with anhydrous magnesium sulfate or a molecular sieve was found to be unsatisfactory because of irreversible adsorption of organics upon the drying agent.

Functional Group Analysis for Carbonyl

Quantitative analysis for carbonyl was conducted using the water-insoluble neutral fraction of the methylene chloride extractable matter. The procedure employed is based on the reaction of aldehydes and ketones with 2,4-dinitrophenylhydrazine (DNPH) to form the corresponding colored hydrazones (A-1,A-2). The color formed is stable for several hours and the molar absorbances of the hydrazones vary so little that the method is well suited to determination of total carbonyl in a mixture.

(A-1) Siggia, S., Quantitative Organic Analysis, John Wiley (1962), p 124.
(A-2) Lappin, G. R., Clark, L. C., Anal. Chem., 23, 541 (1951).

The analyses were conducted in carbonyl-free methanol. This was prepared by refluxing 500 ml of reagent-grade methanol with 2 g of DNPH and 1 ml of concentrated hydrochloric acid for 24 hours. The methanol was then distilled from the mixture, followed by redistillation.

Determinations of unknowns and calibration standards were conducted as follows. In a 10 ml volumetric flask were combined 1 ml of sample in carbonyl free methanol, 1 ml of carbonyl free methanol saturated with DNPH, and one drop of concentrated hydrochloric acid. The flask was stoppered, and heated at 50 C for 30 minutes. After cooling, the reaction mixture was treated with 1 ml of 10 percent potassium hydroxide in 20 percent aqueous methanol (carbonyl free). (In the procedures cited above (A-1,A-2), the addition of 5 ml of potassium hydroxide solution is called for. In the current work, however, it was observed that use of more than 1 ml leads to formation of a precipitate.) Finally, the volume of the mixture was then made up to 10 ml with carbonyl-free methanol. The optical density of the sample at 480 nm was determined with reference to a blank reaction mixture which incorporated 1 ml of carbonyl free methanol in place of the sample.

Calibration experiments were carried out using 5×10^{-2} molar solutions of n-heptaldehyde and 2-heptanone; between 1 and 25 μl of these solutions were added to 1 ml of carbonyl-free methyl alcohol. A linear calibration over the range 10^{-7} to 10^{-6} moles of carbonyl against optical density was thus obtained.

The concentrated methylene chloride extracts used in this determination generally have a concentration of 2 mg/ml. Typically, 250 μl of such solutions contained carbonyl in the range of 10^{-6} - 10^{-7} moles. Analysis of distilled-in-glass methylene chloride (obtained from Burdick and Jackson, Muskegon Michigan) indicated that the solvent is free of appreciable concentrations of carbonyl contaminants. Finally, it should be pointed out that solutions should not exceed 10^{-3} molar in carbonyl in order to prevent precipitation of hydrazones.

Functional Group Analysis for Alcohol

Quantitative analysis for alcohol was conducted using the water-insoluble neutral fraction of the methylene chloride extractable matter. The sample was thoroughly dried before analyses, as described above. The procedure employed is based upon the reaction of alcohols with lithium aluminum hydride to liberate hydrogen. It was adapted from a large-scale procedure in which about 10^{-2} moles of sample is required, and the evolved hydrogen is collected in a gas burette (A-3,A-4). In the modified procedure 10^{-6} moles of sample may be conveniently analyzed, and the evolved hydrogen is measured gas chromatographically.

The apparatus employed consists of a 0.2-ml reaction vessel situated along a capillary gas loop. The vessel is equipped with a screw cap and Teflon-lined injection septum. The apparatus is plumbed to permit gas chromatographic carrier gas to be diverted through the gas loop. Evolved hydrogen is then swept onto the chromatography column and is quantitated using a thermal conductivity detector.

Typically, the analyses were conducted as follows. A saturated solution (~ 10 percent w/w) of lithium aluminum hydride was prepared in sodium-dried tetrahydrofuran, and was stored in a vessel equipped with a Teflon-lined septum cap. 10 μl of reagent was withdrawn

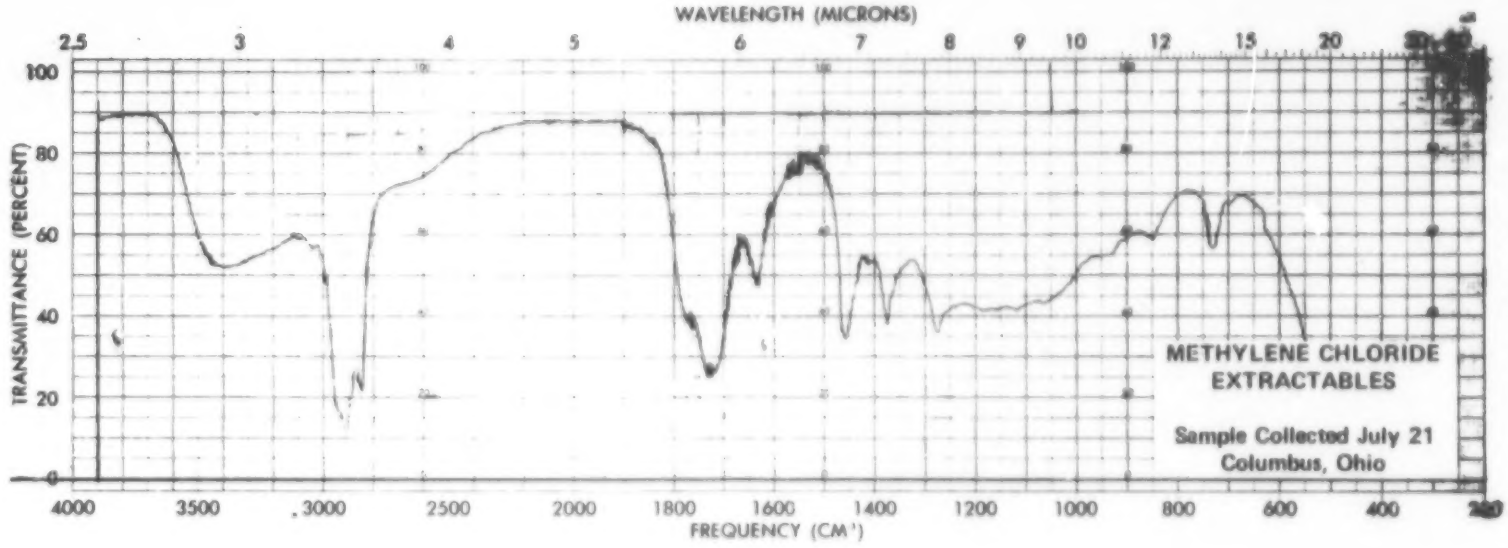
(A-3) Siggia, S., Quantitative Organic Analysis, John Wiley (1963), p 8.
(A-4) Gaylord, N. G., Reduction with Complex Metal Hydrides, Interscience (1956).

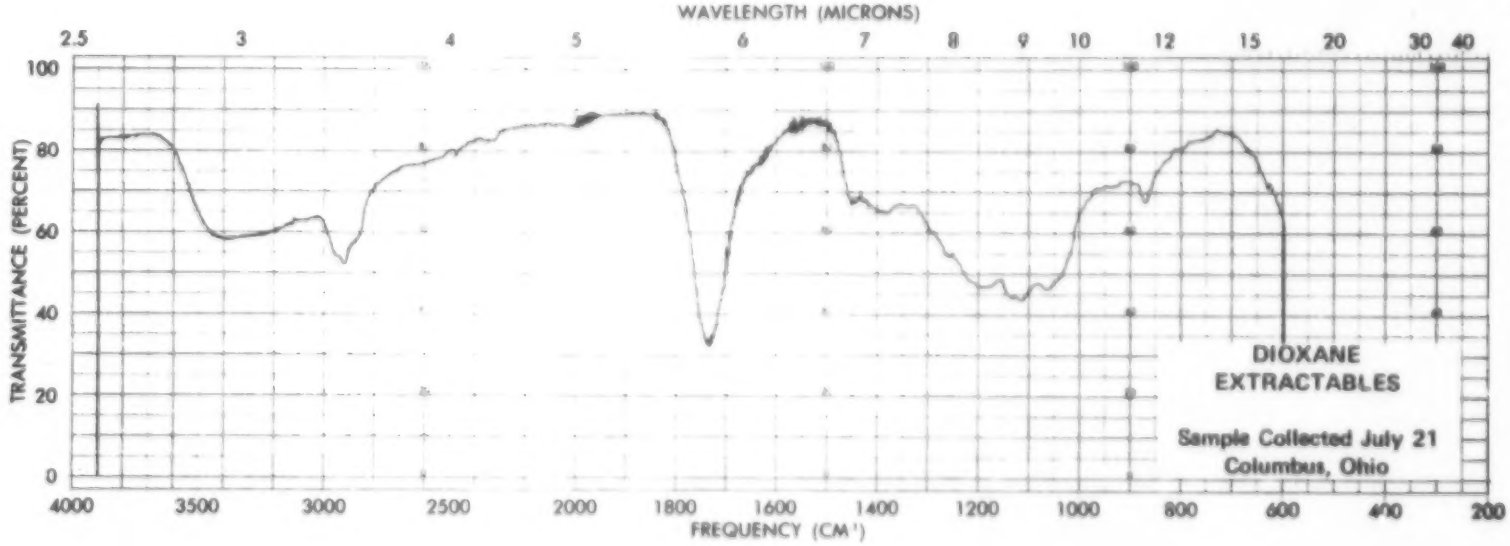
by syringe and injected into the reaction vessel. The solution was repeatedly degassed, by diverting the chromatograph carrier gas through the loop, until hydrogen was no longer present in the injection, or until a very low constant value was achieved. Commercial dry nitrogen was used as the carrier gas, and was further dried by passing through a dry-ice trap.

The test sample was injected into the reaction vessel, and after one minute the evolved hydrogen was swept into the chromatographic column. Longer reaction times and mechanical agitation were without effect. A 1-ft x 2-mm ID silica gel column was used at room temperature. The thermal conductivity detector employed was obtained from a Varian Autoprep chromatograph. Calibration experiments were carried out by injecting 1-10 μl of 0.10 percent or 1.0 percent solutions of cyclohexanol in dry tetrahydrofuran. Typically four replicate analyses were conducted for each calibration standard or unknown. Reproducible results and linear calibrations were observed over the range of 10^{-5} - 10^{-8} moles of alcohol per injection. Typically, a 1-2 mg sample of fractionated particulate extract contained 10^{-6} moles of alcohol. Thus, the sensitivity of the method is entirely adequate. If necessary, additional sensitivity could be achieved by using a thermal conductivity detector of lower thermal capacity.

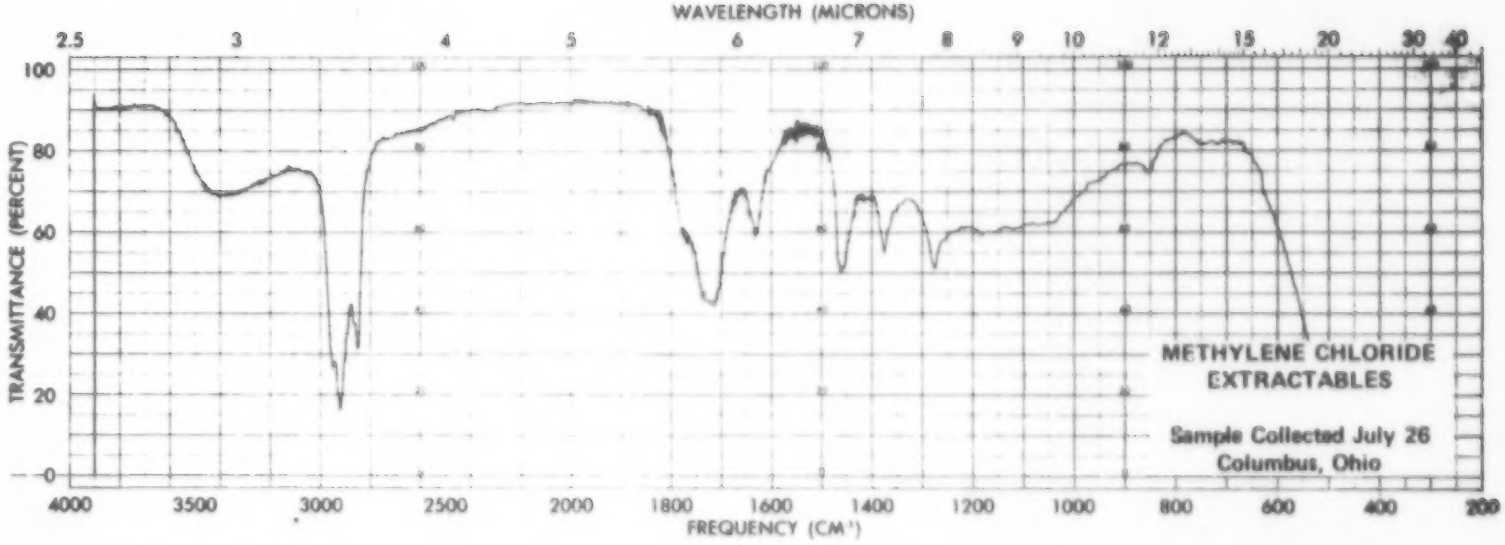
APPENDIX C - i

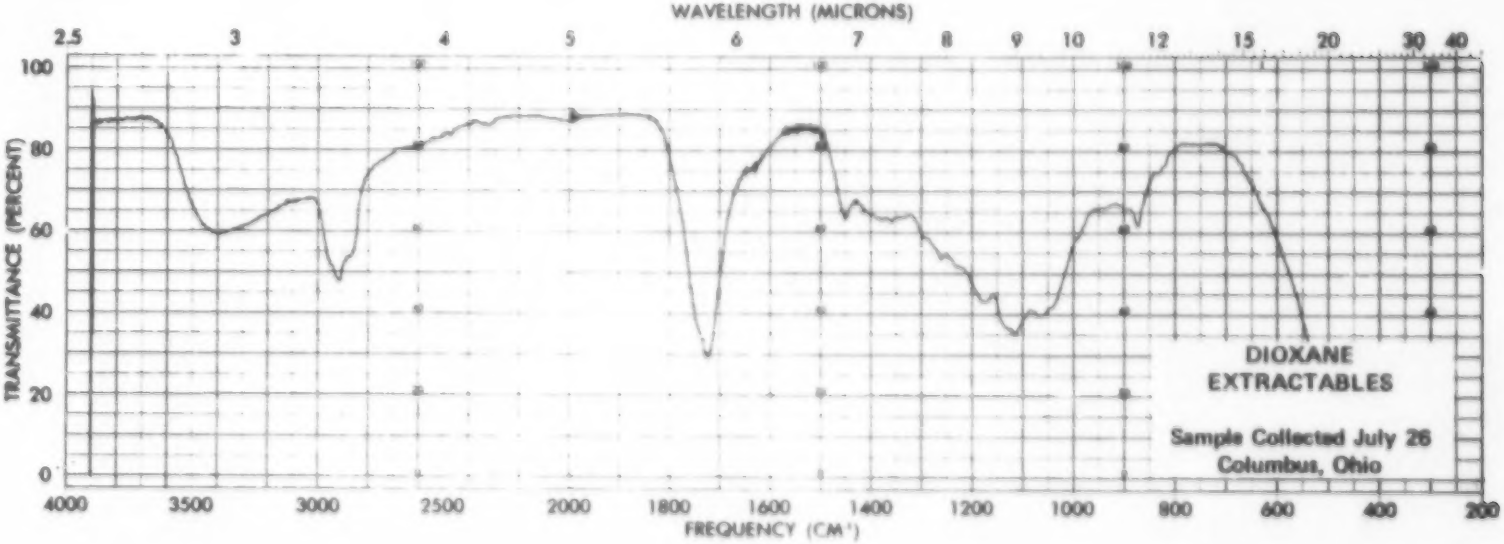
INFRARED SPECTRA OF ORGANIC PARTICULATE

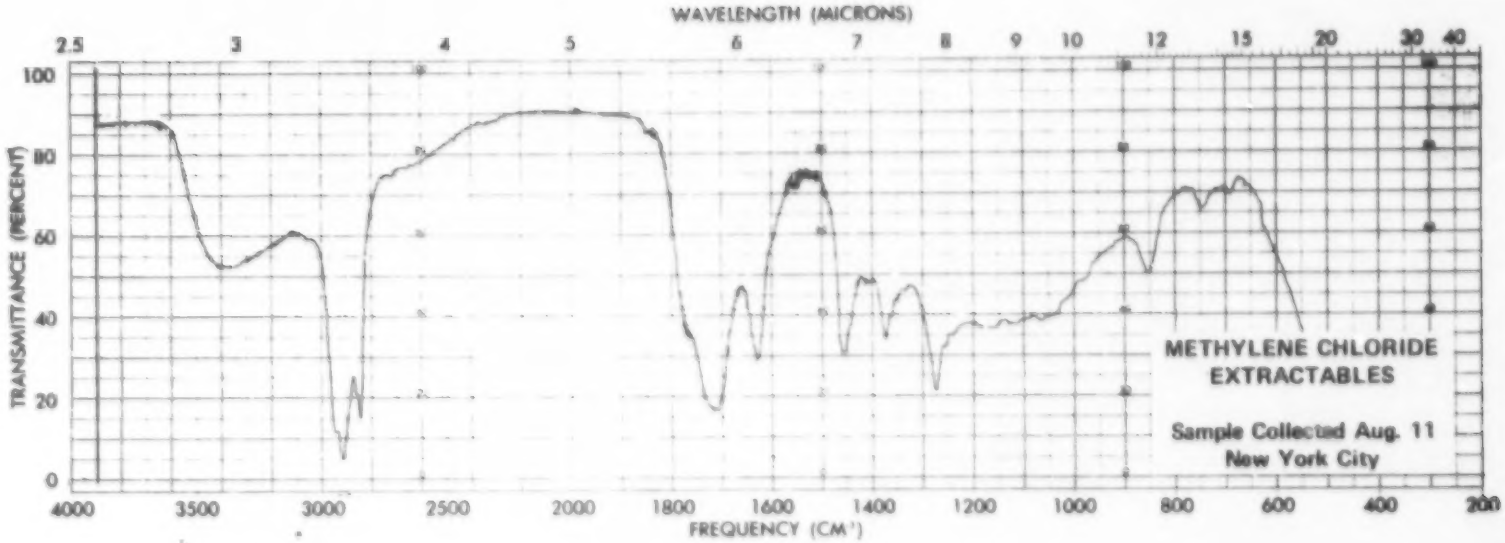

WAVELENGTH (MICRONS)
2.5
3
4
5
6
7
8
9
10
12
15
20
TRANSMITTANCE (PERCENT)
100
80
60
40
20
0
4000
3500
3000
2500
2000
1800
1600
1400
1200
1000
800
600
400
FREQUENCY (CM⁻¹)
METHYLENE CHLORIDE
EXTRACTABLES
Sample Collected July 21
Columbus, Ohio

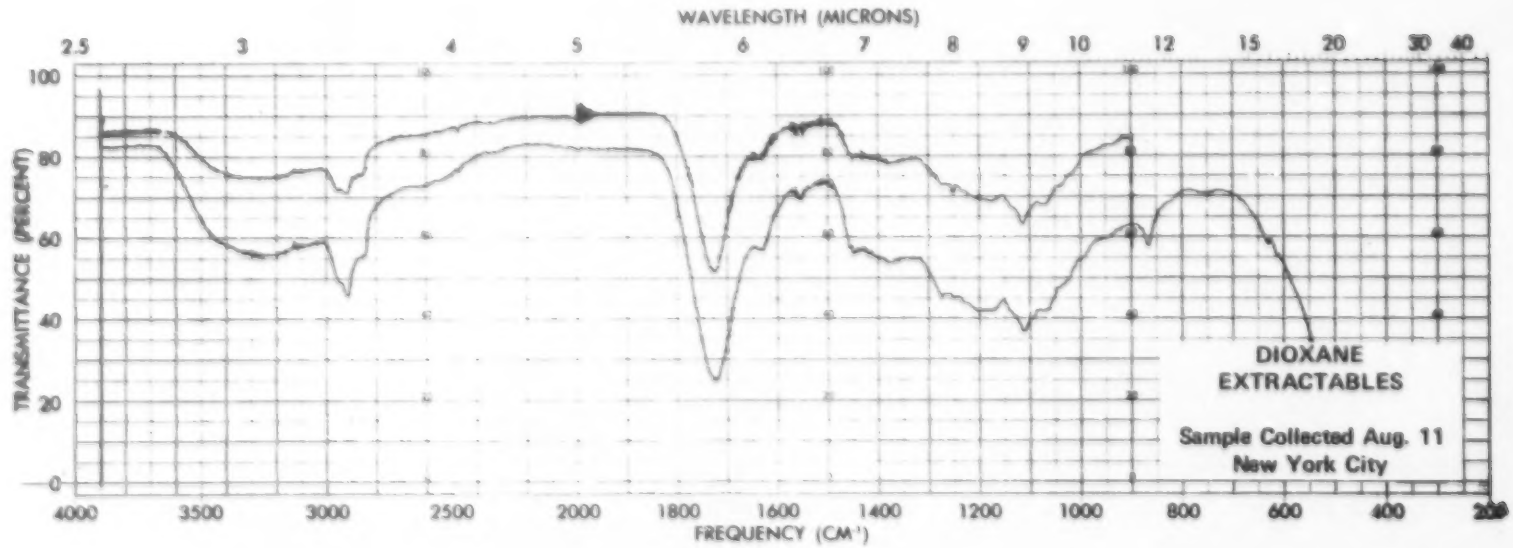

WAVELENGTH (MICRONS)
2.5
3
4
5
6
7
8
9
10
12
15
20
30
40
TRANSMITTANCE (PERCENT)
100
80
60
40
20
0
4000
3500
3000
2500
2000
1800
1600
1400
1200
1000
800
600
400
200
FREQUENCY (CM⁻¹)
DIOXANE
EXTRACTABLES
Sample Collected July 21
Columbus, Ohio

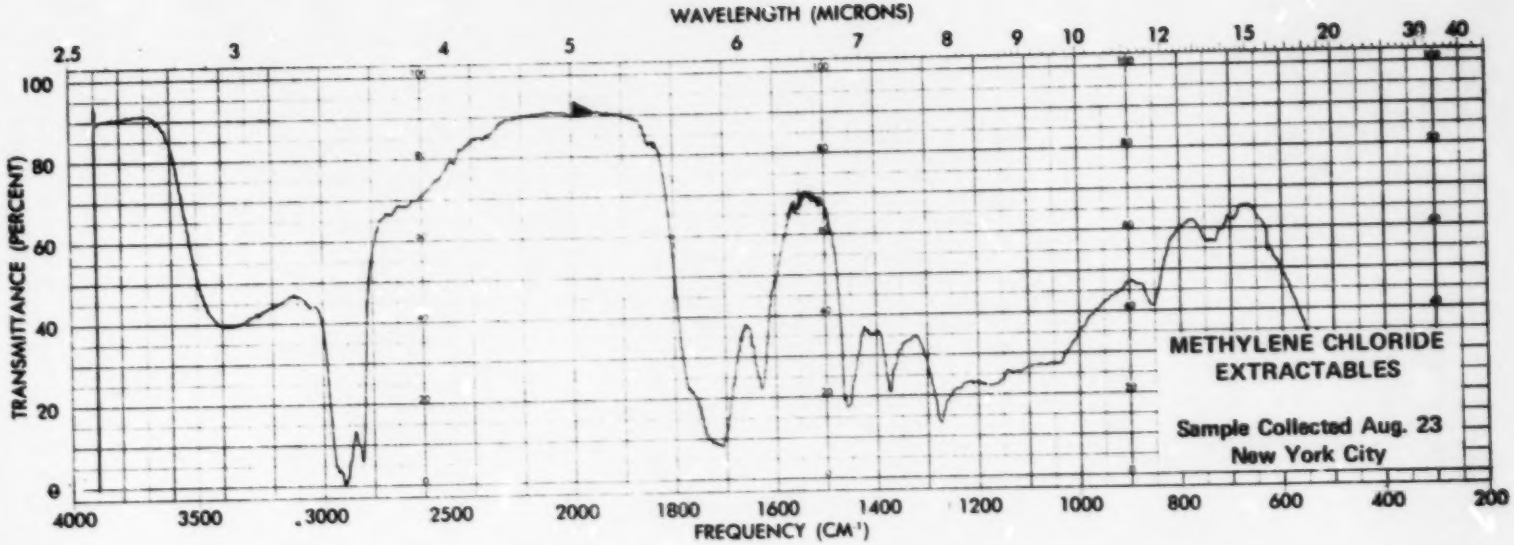
C-1

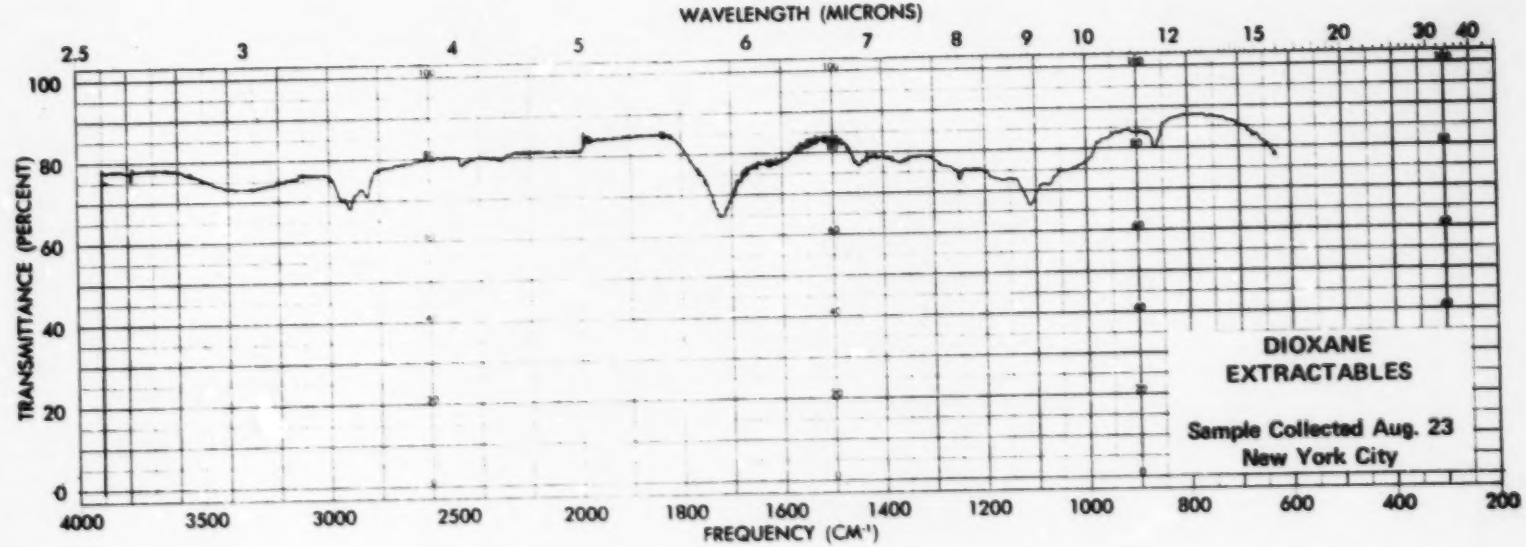

WAVELENGTH (MICRONS)
2.5
3
4
5
6
7
8
9
10
12
15
20
30
40
100
80
60
40
20
0
TRANSMITTANCE (PERCENT)
4000
3500
3000
2500
2000
1800
1600
1400
1200
1000
800
600
400
200
FREQUENCY (CM⁻¹)
METHYLENE CHLORIDE
EXTRACTABLES
Sample Collected July 26
Columbus, Ohio

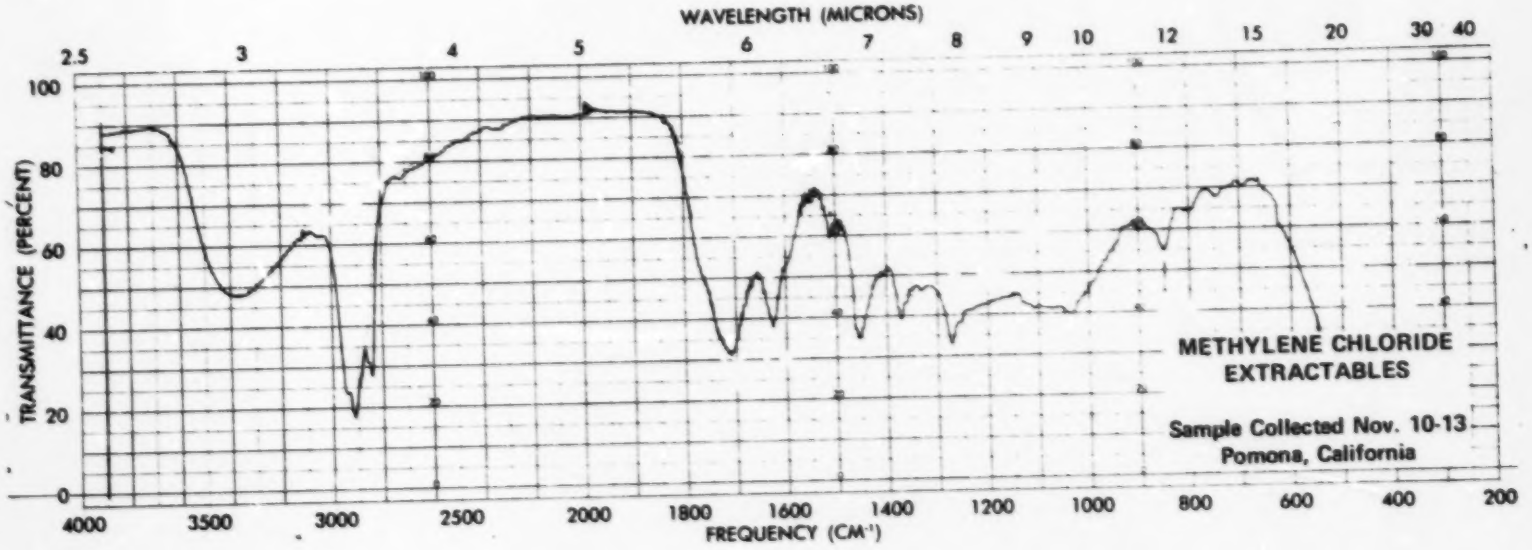

WAVELENGTH (MICRONS)
2.5
3
4
5
6
7
8
9
10
12
15
20
30
40
100
80
60
40
20
0
TRANSMITTANCE (PERCENT)
4000
3500
3000
2500
2000
1800
1600
1400
1200
1000
800
600
400
200
FREQUENCY (CM⁻¹)
DIOXANE
EXTRACTABLES
Sample Collected July 26
Columbus, Ohio

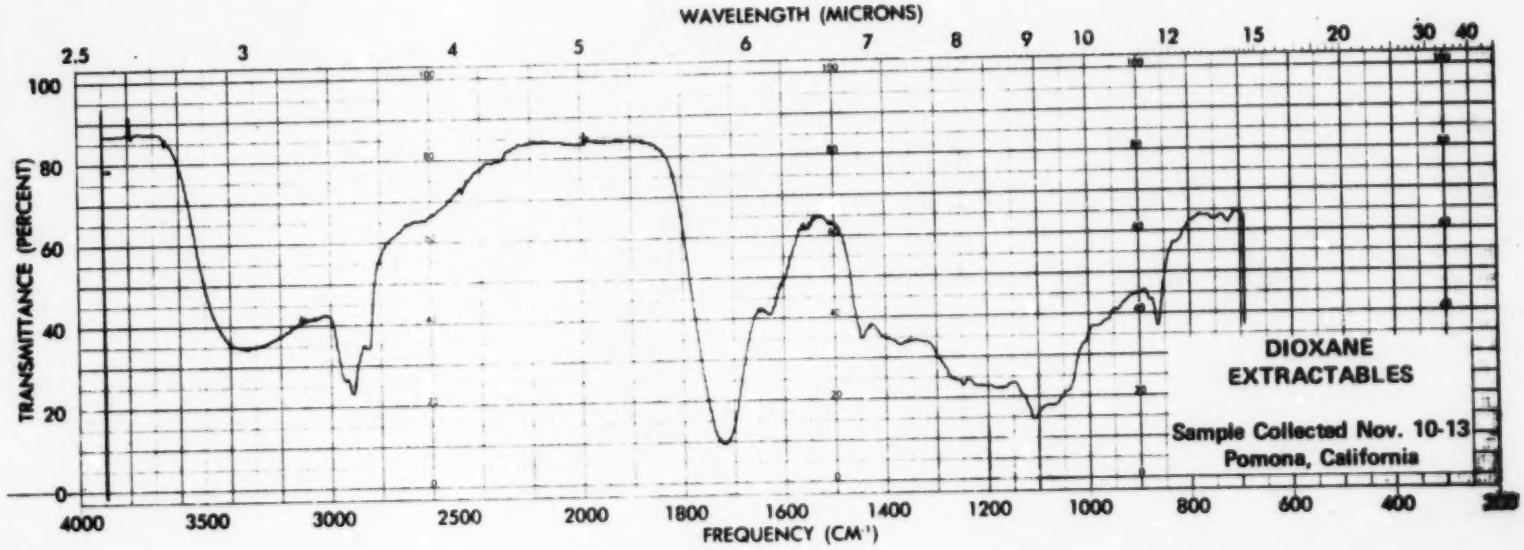

WAVELENGTH (MICRONS)
2.5
3
4
5
6
7
8
9
10
12
15
20
30
40
100
80
60
40
20
0
TRANSMITTANCE (PERCENT)
4000
3500
3000
2500
2000
1800
1600
1400
1200
1000
800
600
400
200
FREQUENCY (CM⁻¹)
METHYLENE CHLORIDE
EXTRACTABLES
Sample Collected Aug. 11
New York City

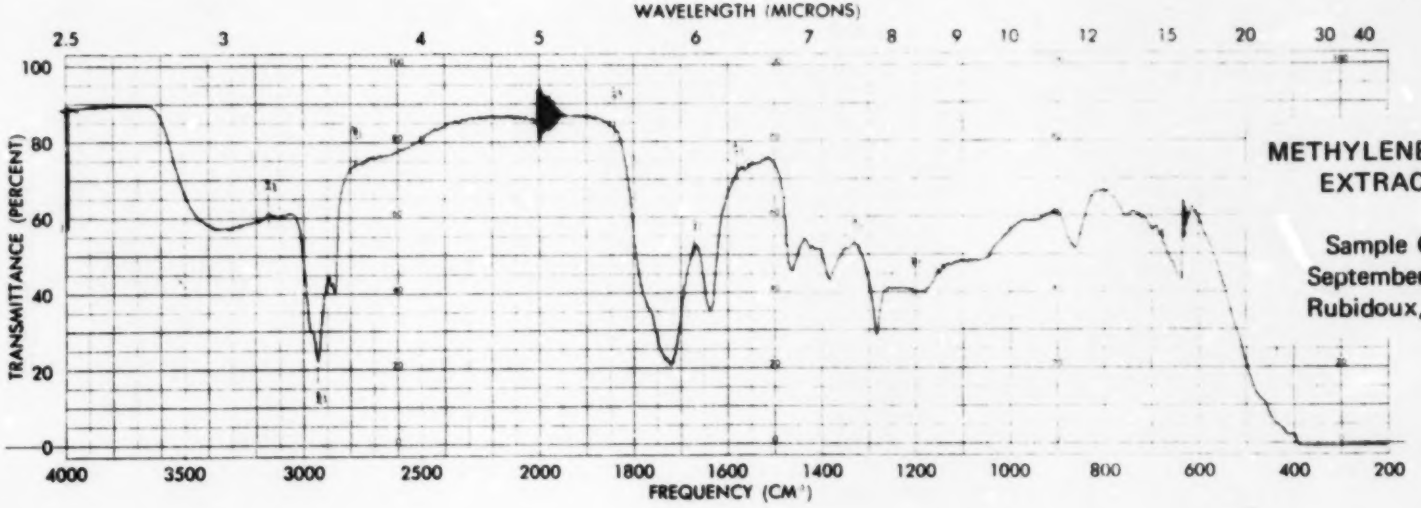

WAVELENGTH (MICRONS)
2.5
3
4
5
6
7
8
9
10
12
15
20
30
40
100
80
60
40
20
0
TRANSMITTANCE (PERCENT)
4000
3500
3000
2500
2000
1800
1600
1400
1200
1000
800
600
400
FREQUENCY (CM⁻¹)
DIOXANE
EXTRACTABLES
Sample Collected Aug. 11
New York City


WAVELENGTH (MICRONS)
2.5
3
4
5
6
7
8
9
10
12
15
20
30
40
TRANSMITTANCE (PERCENT)
100
80
60
40
20
0
4000
3500
3000
2500
2000
1800
1600
1400
1200
1000
800
600
400
200
FREQUENCY (CM⁻¹)
METHYLENE CHLORIDE
EXTRACTABLES
Sample Collected Aug. 23
New York City


WAVELENGTH (MICRONS)
2.5
3
4
5
6
7
8
9
10
12
15
20
30
40
TRANSMITTANCE (PERCENT)
100
80
60
40
20
0
4000
3500
3000
2500
2000
1800
1600
1400
1200
1000
800
600
400
200
FREQUENCY (CM⁻¹)
DIOXANE
EXTRACTABLES
Sample Collected Aug. 23
New York City

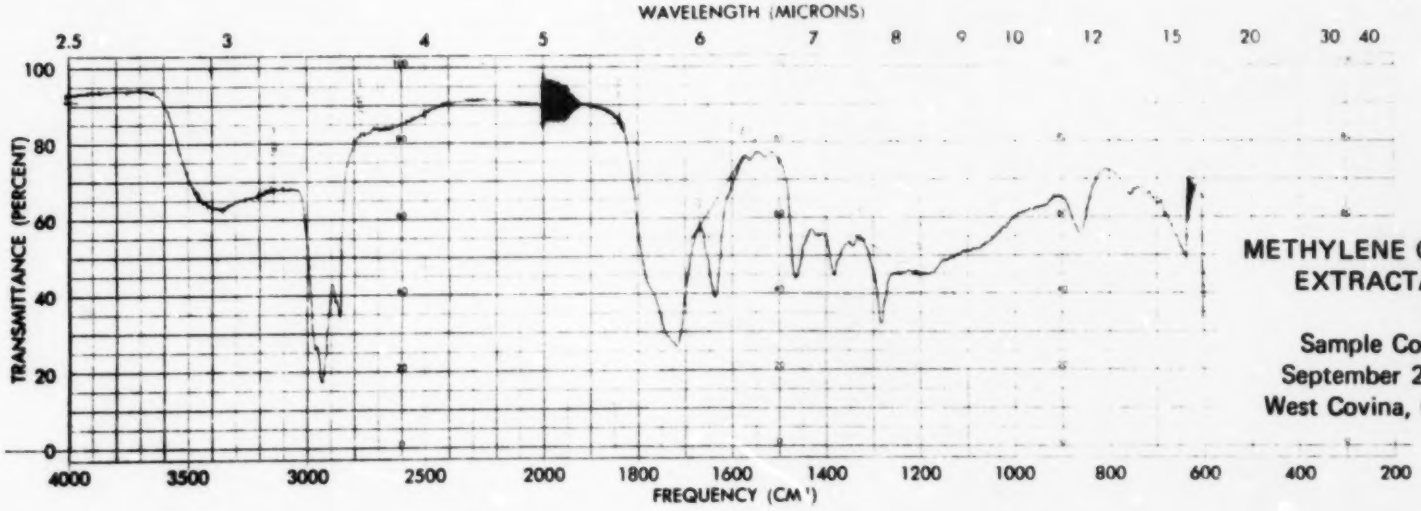

WAVELENGTH (MICRONS)
2.5
3
4
5
6
7
8
9
10
12
15
20
30
40
100
80
60
40
20
0
TRANSMITTANCE (PERCENT)
4000
3500
3000
2500
2000
1800
1600
1400
1200
1000
800
600
400
200
FREQUENCY (CM⁻¹)
METHYLENE CHLORIDE
EXTRACTABLES
Sample Collected Nov. 10-13
Pomona, California


WAVELENGTH (MICRONS)
2.5
3
4
5
6
7
8
9
10
12
15
20
30
40
100
80
60
40
20
0
TRANSMITTANCE (PERCENT)
4000
3500
3000
2500
2000
1800
1600
1400
1200
1000
800
600
400
FREQUENCY (CM⁻¹)
DIOXANE
EXTRACTABLES
Sample Collected Nov. 10-13
Pomona, California



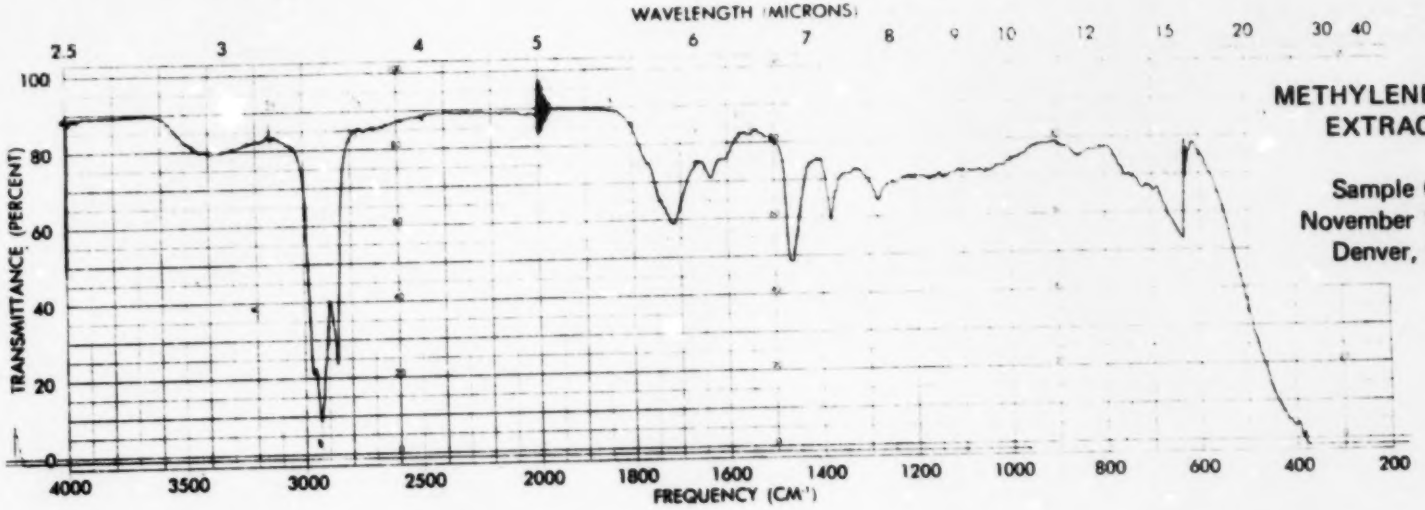
METHYLENE CHLORIDE EXTRACTABLES

Sample Collected
September 21, 1973
Rubidoux, California



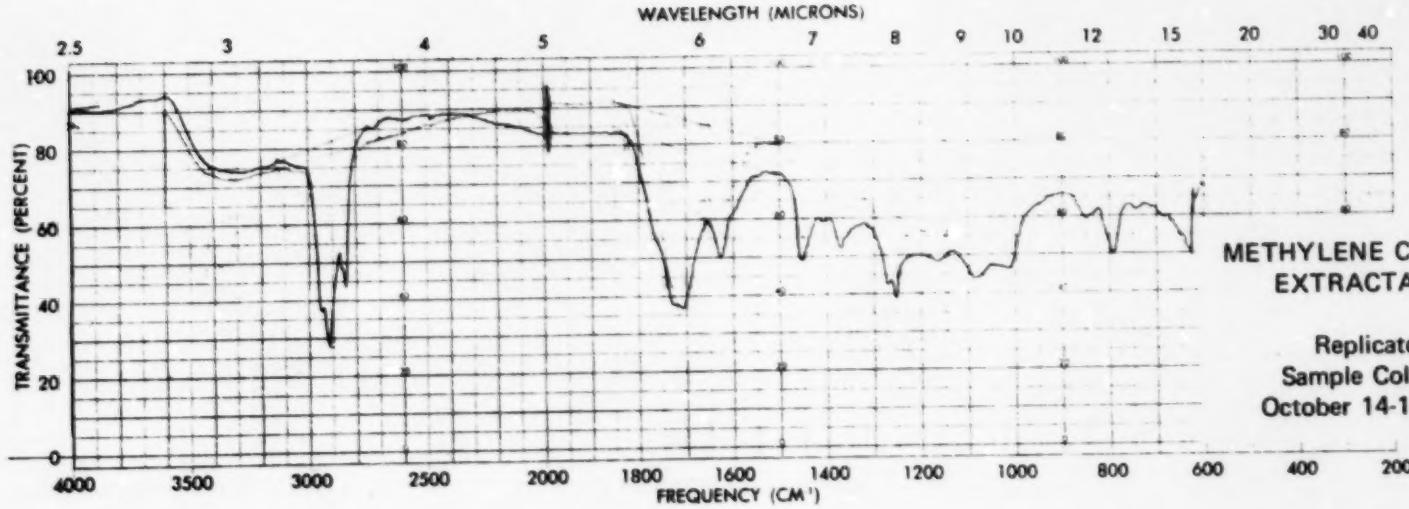
METHYLENE CHLORIDE EXTRACTABLES

Sample Collected
September 21, 1973
West Covina, California



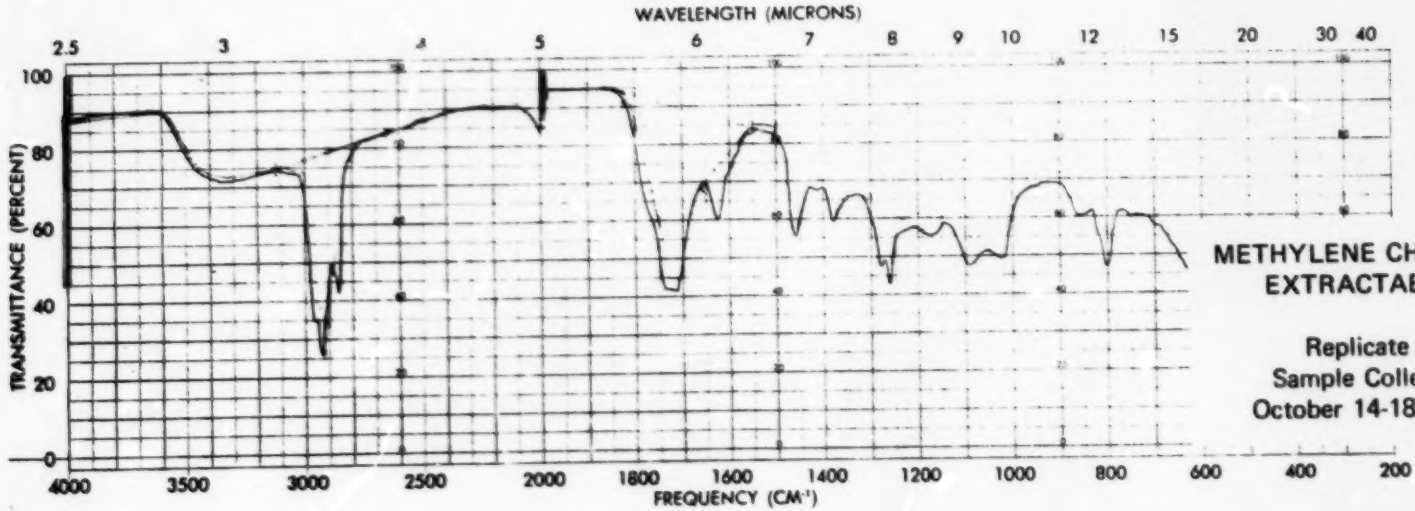
METHYLENE CHLORIDE EXTRACTABLES

Sample Collected
November 16-17, 1973
Denver, Colorado



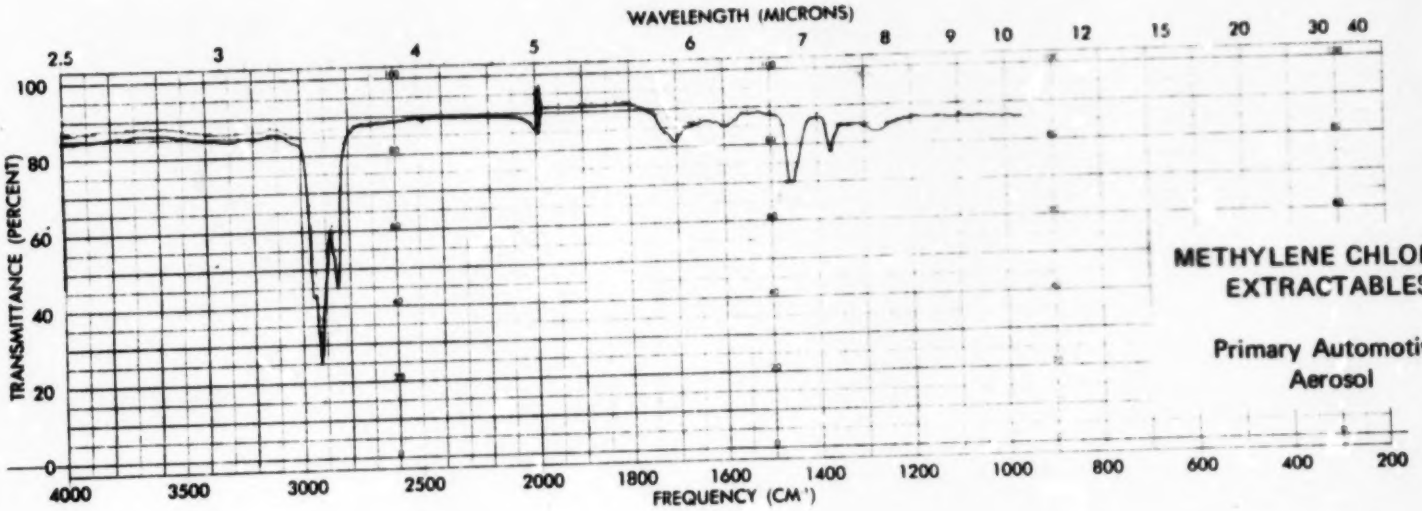
METHYLENE CHLORIDE EXTRACTABLES

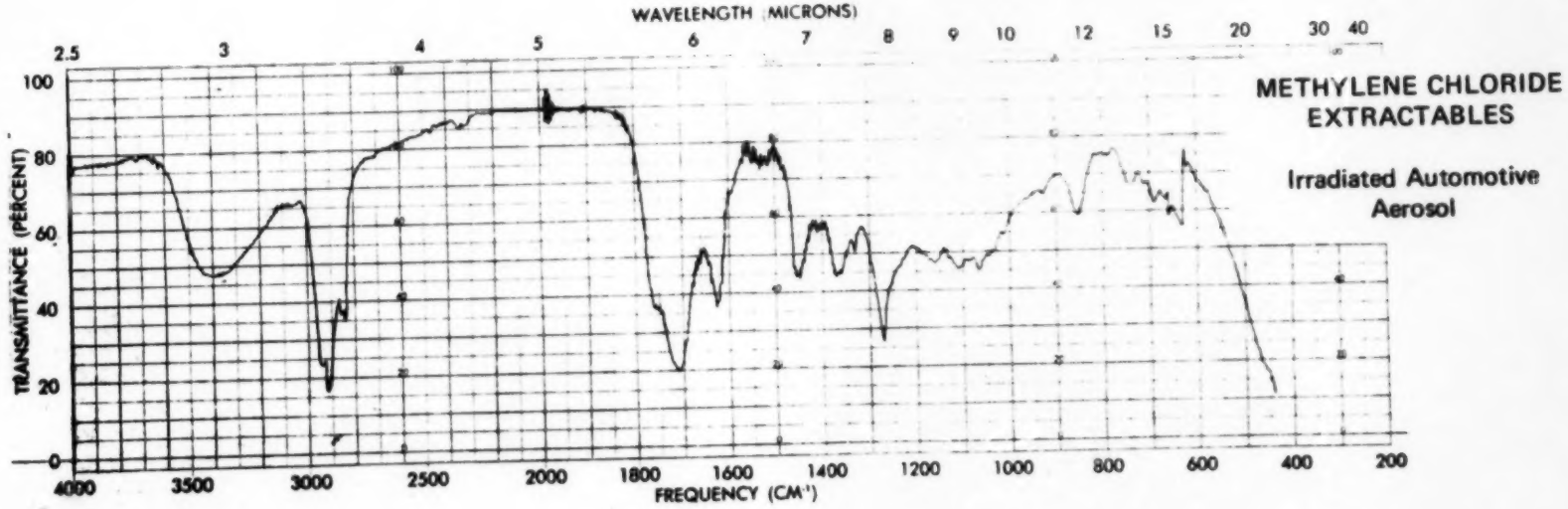
Replicate A
Sample Collected
October 14-18, 1972



METHYLENE CHLORIDE EXTRACTABLES

Replicate B
Sample Collected
October 14-18, 1972


WAVELENGTH (MICRONS)
2.5
3
4
5
6
7
8
9
10
12
15
20
30
40
100
80
60
40
20
0
TRANSMITTANCE (PERCENT)
4000
3500
3000
2500
2000
1800
1600
1400
1200
1000
800
600
400
200
FREQUENCY (CM⁻¹)
METHYLENE CHLORIDE EXTRACTABLES
Primary Automotive Aerosol


WAVELENGTH (MICRONS)
2.5
3
4
5
6
7
8
9
10
12
15
20
30
40
100
80
60
40
20
0
TRANSMITTANCE (PERCENT)
4000
3500
3000
2500
2000
1800
1600
1400
1200
1000
800
600
400
200
FREQUENCY (CM⁻¹)
METHYLENE CHLORIDE EXTRACTABLES
Irradiated Automotive Aerosol